

Access DB#

71413

# SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: N A WALICKA Examiner #: 78201 Date: July 19, 2002  
 Art Unit: 1652 Phone Number 305-7270 Serial Number: 04/657,986  
 Mail Box and Bldg/Room Location: 10000 Results Format Preferred (circle): PAPER DISK E-MAIL  
10001

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Inhibitors of SERINE PROTEASE ACTIVITY

Inventors (please provide full names): Madison et al

Earliest Priority Filing Date: 09/08/2000

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search

SEQ ID NO: 2 , ASAP please

Thank you

Infoclick

(1010)

RECEIVED  
 JUL 19 2002

Point of Contact:  
 Toby Port  
 Technical Info. Specialist  
 CM1 6A04  
 703-308-3534

## STAFF USE ONLY

Searcher: \_\_\_\_\_

Searcher Phone #: \_\_\_\_\_

Searcher Location: \_\_\_\_\_

Date Searcher Picked Up: 7/22

Date Completed: 7/23

Searcher Prep & Review Time: 10

Clerical Prep Time: \_\_\_\_\_

Online Time: 11

## Type of Search

NA Sequence (#) \_\_\_\_\_

AA Sequence (#) 1

Structure (#) \_\_\_\_\_

Bibliographic \_\_\_\_\_

Litigation \_\_\_\_\_

Fulltext \_\_\_\_\_

Patent Family \_\_\_\_\_

Other \_\_\_\_\_

## Vendors and cost where applicable

STN \_\_\_\_\_

Dialog \_\_\_\_\_

Questel/Orbit \_\_\_\_\_

Dr. Link \_\_\_\_\_

Lexis/Nexis \_\_\_\_\_

Sequence Systems CS

WWW/Internet \_\_\_\_\_

Other (specify) \_\_\_\_\_

GenCore version 4.5.  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:07:54 ; Search time 29.61 Seconds  
(without alignments)  
904.045 Million cell updates/sec

Title: US-09-657-986b-2

Perfect score: 1319

Sequence: 1 VVGFTDADEGEWPCVSLHA.....PGVYTRLPFRDWIKENTGV 241

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1319	100.0	241	AAE06936	Human membrane-type
2	1319	100.0	683	AA19551	Human matriptase
3	1319	100.0	762	AA190284	Human peptidase, H
4	1319	100.0	851	AB111428	Human membrane-type
5	1319	100.0	851	AA25628	Human protein sequ
6	1319	100.0	855	AA06671	Tumour antigen der
7	1319	100.0	855	AA19552	Human matriptase
8	1319	100.0	855	AAE06930	Human membrane-type
9	1319	100.0	855	AA198500	Human TADG-15. HO
10	1319	100.0	855	AA198500	Human membrane-type
11	1315	99.7	241	AAW22987	Human serine prote

12	1227	93.0	932	22	ABG21442	Novel human diagno
13	1100	83.4	902	22	AAE06936	Marine epithin. M
14	643	48.7	129	22	ABG21439	Novel human diagno
15	576.5	43.7	233	22	AAE01942	Human endotheliase
16	576.5	43.7	422	21	AA194708	Human DESCI protei
17	576.5	43.7	423	21	AA199414	Human PRO1461 (UNQ
18	576.5	43.7	423	22	AAU29183	Human PRO polytype
19	576.5	43.7	423	22	AAU01344	Human TANGO 361 am
20	576.5	43.7	423	22	AAU01399	Human TANGO 361, v
21	576.5	43.7	423	22	AAU01400	Human TANGO 361, v
22	576.5	43.7	423	22	AAU01401	Human TANGO 361, v
23	576.5	43.7	423	22	AAU01402	Human TANGO 361, v
24	576.5	43.7	423	22	AA187578	Human PRO1461. HO
25	576.5	43.7	423	22	AA186613	Protein of the inv
26	563.5	42.7	422	22	AAE01946	Human endotheliase
27	562.5	42.6	422	21	AA194709	Human DESCI protei
28	561.5	42.6	235	22	AAE06932	Human membrane-type
29	561.5	42.6	658	22	AAE06934	Human membrane-type
30	561.5	42.6	802	22	AAE06933	Human membrane-type
31	559.5	42.4	802	20	AA141710	Human PRO618 prote
32	559.5	42.4	802	21	AA144266	Human PRO618 (UNQ3
33	559.5	42.4	802	21	AA124052	Human PRO618 prote
34	547.5	41.5	1019	22	AAE06940	Human enterokinase
35	544.5	41.3	798	15	AA157283	Bovine enterokinas
36	540	40.9	238	22	AAE06937	Novel human diagno
37	506	38.4	145	21	AA190298	Human peptidase, H
38	501.5	38.0	250	22	AB165923	Drosophila melanog
39	499	37.8	492	21	AA136901	Human TMPRSS2 prot
40	499	37.8	492	21	AA157280	Ovrl15 homolog pro
41	499	37.8	492	22	AA169939	Human transmembran
42	499	37.8	492	22	AAE06943	Human serine prote
43	499	37.8	492	22	AA101294	Human transmembran
44	497.5	37.7	232	17	AA189430	Trypsin-like enzym
45	497.5	37.7	418	17	AA189435	Trypsin-like enzym

ALIGNMENTS

RESULT 1

AAE06936

ID AAE06936 standard; Protein; 241 AA.

AC AAE06936;

DT 16-OCT-2001 (first entry)

DE Human membrane-type serine protease (MTSP) 1 protease domain.

DE Human; transmembrane serine protease; membrane-type serine protease;  
KW MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;  
KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;  
KW matriptase.

OS Homo sapiens.

PN WO200157194-A2.

PD 09-AUG-2001.

PF 02-FEB-2001; 2001WO-US03471.

PR 03-FEB-2000; 2000US-0179982.

PR 18-FEB-2000; 2000US-0183542.

PR 22-JUN-2000; 2000US-0213124.

PR 26-JUL-2000; 2000US-0220970.

PR 08-SEP-2000; 2000US-0657986.

PR 22-SEP-2000; 2000US-0234640.

XX (CORV-) CORVAS INT INC.

XX Madison EL, Ong EO, Yeh J;

```

DR WPI; 2001-488877/53.
DR N-PSDB; AAD13155.
XX
XX Novel single chain polypeptide comprising protease domain of type-II
XX membrane-type serine protease or its catalytically active portion
XX useful for treating and preventing cancer and tumor
XX
XX Example 5; Page 227; 256pp; English.
XX
XX The invention relates to transmembrane serine proteases and their
XX corresponding nucleotides and the protease domain of a type-II
XX membrane-type serine protease (MTSP). MTSP is useful for identifying
XX compounds that modulate or inhibit its proteolytic activity and for
XX formulating a medicament for treating neoplastic disease. MTSP and
XX its corresponding nucleotides are useful in preventing or treating
XX tumors or cancers such as lung carcinoma, colon adenocarcinoma and
XX ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
XX is useful as a diagnostic marker for tumor development, growth and/or
XX progression and as immunogens to generate antibodies that specifically
XX bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
XX gene therapy. The present sequence is protease domain of human MTSP1
XX protein (also called matriptase).
SQ Sequence 241 AA;

Query Match 100.0%; Score 1319; DB 22; Length 241;
Best Local Similarity 100.0%; Pred. NC. 1.1e-98;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTDADGEGWPQVSLHALGQGHICGASLSPNWLVSAAHCYIDDRGFYSDPTQWTA 60
Db 1 VVGTDADGEGWPQVSLHALGQGHICGASLSPNWLVSAAHCYIDDRGFYSDPTQWTA 60
QY 61 FLGLHDQSQSAQCYQERLKRITISHPPNDFDYDIALLEKPAEYSVMWRPCLPD 120
Db 61 FLGLHDQSQSAQCYQERLKRITISHPPNDFDYDIALLEKPAEYSVMWRPCLPD 120
QY 121 ASHVFPAGKAIWVWGHTQYGGTGALILQKEIRVINQTCENLPOQITPRMVCVGL 180
Db 121 ASHVFPAGKAIWVWGHTQYGGTGALILQKEIRVINQTCENLPOQITPRMVCVGL 180
QY 181 SGGVDSQSGSGPLSSVEADGRIFQAGVSVWGDCAGRNKPGVYTRLPFRWIKENTG 240
Db 181 SGGVDSQSGSGPLSSVEADGRIFQAGVSVWGDCAGRNKPGVYTRLPFRWIKENTG 240
QY 241 V 241
Db 241 V 241

RESULT 2
AAB19551
ID AAB19551 standard; Protein; 683 AA.
XX
XX AAB19551;
XX
XX 22-JAN-2001 (first entry)
XX
XX Human matriptase (truncated form).
XX
XX Matriptase; serine protease; human; breast cancer; pre-malignancy;
XX actinic keratosis; leukoplakia; Barrett's epithelium;
XX colunar metaplasia; ulcerative colitis; Bowenoid papulosis;
XX adenomatous colorectal polyp; Ovarat erythroplasia;
XX vulvar intraepithelial neoplasia; tumour; metastasis; therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..31
XX Protein 32..683
  
```

---

```

/label= Mature_protein
42..155
/note= "complement subcomponent 1r and 1s (Clr/s)
region"
168..268
/note= "complement subcomponent 1r and 1s (Clr/s)
region"
280..314
/note= "LDL-receptor type region"
315..351
/note= "LDL-receptor type region"
352..387
/note= "LDL-receptor type region"
394..430
/note= "LDL-receptor type region"
432..683
/label= Catalytic_domain
130
/modified-site
130
/modified-site
313
/modified-site
313
/modified-site
484
/active-site
484
/active-site
539
/active-site
539
/active-site
633
/active-site
633
/active-site
77..79
/peptide
77..79
/peptide
469..485
/disulfide-bond
469..485
/disulfide-bond
604..618
/disulfide-bond
604..618
/disulfide-bond
629..658
/cleavage-site
629..658
/cleavage-site
442..446
/conserved
442..446
/conserved
/note= "conserved proteolytic activation site"
W0200053232-A1.
14-SEP-2000.
10-MAR-2000; 2000WO-US06111.
12-MAR-1999; 99US-0124006.
(GEOU ) UNIV GEORGETOWN.
Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;
WPI; 2000-594268/56.
N-PSDB; AAA88492.
Treating malignancies, premalignant and pathologic conditions in a
subject, comprises administering matriptase modulating agent
Claim 14; Fig 9; 116pp; English.
The present sequence is that of the truncated form of human
matriptase, a trypsin-like protease, as deduced from cDNA (see
AAA88492) obtained from human breast cancer cell cDNA by RT-PCR.
The full-length form (see AAB19552) has an additional 172 N-terminal
amino acids. Either form can be produced in transformed or
transfected cells using a claimed method. The zymogen (inactive)
form of matriptase is a single-chain protein. The active 2-chain
form strongly interacts with fragments of a Kunitz-type serine
protease inhibitor (hepatocyte growth factor activator inhibitor,
HAI-1) to form SDS-stable complexes. In breast cancer cells,
matriptase is present mainly as the uncomplexed form. Only the
complexed matriptase is detected in human milk. The invention is
directed to a method of detecting a malignancy or a pre-malignant
lesion in breast or other tissue by detecting the presence of
single- or 2-chain forms of matriptase in the tissue. The
  
```

CC object is to inhibit tumour onset, tumour growth and metastasis.  
 CC Malignancies and pre-malignant conditions characterised by  
 CC expression of the zymogen or activated form of matrilptase are  
 CC created by administering an inhibitor of matrilptase, especially a  
 CC Bowman-Birk inhibitor. The pre-malignant condition is atypical  
 CC ductal hyperplasia of the breast, actinic keratosis, leukoplakia,  
 CC Barrett's epithelium of the esophagus, ulcerative colitis,  
 CC adenomatous colorectal polyps, erythroplasia of the Queyrat,  
 CC Bowen's disease, Bowenoid papulosis, vulvar intraepithelial  
 CC neoplasia or dysplastic changes to the cervix. The invention also  
 CC provides methods for in vivo or in vitro diagnosis of malignancy  
 CC or pre-malignant lesion, and methods of identifying matrilptase  
 CC modulators, including activators and inhibitors.

XX Sequence 683 AA;

Query Match 100.0%; Score 1319; DB 21; Length 683;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-98;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVGTDADGEWPMQVSHALGQGHICGASLISPNMLVSAHCYIDDRGFYSDPTQWTA 60  
 Db 443 VVGTDADGEWPMQVSHALGQGHICGASLISPNMLVSAHCYIDDRGFYSDPTQWTA 502  
 Qy 61 FLGLHDSQRSAPGVQERRLKRISHPFFNDFTDIALLELEKPAEYSSMVRPCLPD 120  
 Db 503 FLGLHDSQRSAPGVQERRLKRISHPFFNDFTDIALLELEKPAEYSSMVRPCLPD 562  
 Qy 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVNOTTCENLLPQGITPRMVCVGL 180  
 Db 563 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVNOTTCENLLPQGITPRMVCVGL 622  
 Qy 181 SGGVDSQCGSGGPLSSVEADGRIFQAGVYVSWDGGCAQRNKPQVYTRLPFRDWIKENTG 240  
 Db 623 SGGVDSQCGSGGPLSSVEADGRIFQAGVYVSWDGGCAQRNKPQVYTRLPFRDWIKENTG 682  
 Qy 241 V 241  
 Db 683 v 683

RESULT 3  
 RAY90284

ID AAY90284 standard; Protein; 762 AA.

XX AC AAY90284;

XX DT 24-OCT-2000 (first entry)

XX DE Human peptidase, HPEP-1 protein sequence.

XX KW Human; peptidase; cell proliferative disorder; arteriosclerosis;  
 KW psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease;  
 KW inflammatory disorder; AIDS; anaemia; allergy; asthma; atherosclerosis;  
 KW Grave's disease; multiple sclerosis; scleroderma; infection; diabetes;  
 KW metabolic disorder; Addison's disease; cystic fibrosis; diagnosis;  
 KW glycogen storage disease; obesity; therapy; HPEP-1.

XX OS Homo sapiens.

XX PN WO200042201-A2.

XX PD 20-JUL-2000.

XX PF 11-JAN-2000; 2000WO-US00641.

XX PR 11-JAN-1999; 99US-0172247.

XX PR 03-MAY-1999; 99US-0132253.

XX PR 27-MAY-1999; 99US-0136653.

XX PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Hillman JL, Tang YT, Azimzai Y, Baughn MR, Lal P;  
 PI Yue H, Lu DAM;  
 XX WPI; 2000-482832/42.  
 DR N-PSDB; AAA37657.  
 XX An isolated polypeptide for diagnosis, prevention and treatment of  
 PT cell proliferative, autoimmune/ inflammatory and metabolic disorders  
 PT comprises a sequence encoding a human peptidase -  
 XX Claim 2; Page 91-93; 13lpp; English.

CC This sequence represents a human peptidase, designated HPEP-1. The  
 CC invention relates to 18 human peptidases designated HPEP-1 to HPEP-18,  
 CC respectively. The peptidases can be used for treating a disease or  
 CC condition associated with decreased expression or over expression of  
 CC functional human peptidases. The diseases that can be diagnosed,  
 CC prevented and treated include cell proliferative disorders (such as  
 CC arteriosclerosis, psoriasis, myelofibrosis, and cancers),  
 CC autoimmune/inflammatory disorders (such as AIDS, anaemia, allergies,  
 CC Crohn's disease, asthma, atherosclerosis, Grave's disease, multiple  
 CC sclerosis, and scleroderma), infections, and metabolic disorders (such as  
 CC Addison's disease, diabetes, cystic fibrosis, glycogen storage diseases  
 CC and obesity).

XX Sequence 762 AA;

Query Match 100.0%; Score 1319; DB 21; Length 762;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-98;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVGTDADGEWPMQVSHALGQGHICGASLISPNMLVSAHCYIDDRGFYSDPTQWTA 60  
 Db 522 VVGTDADGEWPMQVSHALGQGHICGASLISPNMLVSAHCYIDDRGFYSDPTQWTA 581  
 Qy 61 FLGLHDSQRSAPGVQERRLKRISHPFFNDFTDIALLELEKPAEYSSMVRPCLPD 120  
 Db 582 FLGLHDSQRSAPGVQERRLKRISHPFFNDFTDIALLELEKPAEYSSMVRPCLPD 641  
 Qy 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVNOTTCENLLPQGITPRMVCVGL 180  
 Db 642 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVNOTTCENLLPQGITPRMVCVGL 701  
 Qy 181 SGGVDSQCGSGGPLSSVEADGRIFQAGVYVSWDGGCAQRNKPQVYTRLPFRDWIKENTG 240  
 Db 702 SGGVDSQCGSGGPLSSVEADGRIFQAGVYVSWDGGCAQRNKPQVYTRLPFRDWIKENTG 761  
 Qy 241 V 241  
 Db 762 v 762

RESULT 4

ABBI1428

ID ABBI1428 standard; peptidase; 851 AA.

XX AC ABBI1428;

XX DT 11-JAN-2002 (first entry)

XX DE Human membrane-type Ser Kinase homologue, SEQ ID NO:1798.

XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;

KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 XX antifungal; vulnerary; antiulcer.  
 CS Homo sapiens.  
 XX  
 PN WO200157188-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US03800.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-457740/49.  
 DR N-PSDB; ABA08672.  
 XX  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 XX Claim 20; Page 188; 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 XX  
 XX Sequence 851 AA;

Query Match 100.08; Score 1319; DB 22; Length 851;  
 Best Local Similarity 100.08; Pred. No. 3.8e-98;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTDADGEWPQVSHALGQGHICGASLISPNLWLSAAHCVYIDDRGRYSPTQWTA 60  
 DB 611 VVGGDADGEWPQVSHALGQGHICGASLISPNLWLSAAHCVYIDDRGRYSPTQWTA 670  
 QY 61 FLGLHDQSORSGAPGVQERRLRKRIISHPEFNDFTFYDIALLEKPAEYSSMVRPICLPD 120  
 DB 671 FLGLHDQSORSGAPGVQERRLRKRIISHPEFNDFTFYDIALLEKPAEYSSMVRPICLPD 730  
 QY 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINTTCENLLPQQITPRMVCYGL 180  
 DB 731 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINTTCENLLPQQITPRMVCYGL 790  
 QY 181 SGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENTG 240  
 DB 791 SGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENTG 850  
 QY 241 V 241  
 DB 851 V 851  
 RESULT 5  
 AAM25628  
 ID AAM25628 standard; Protein; 851 AA.  
 XX  
 AC AAM25628;  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:1143.  
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;  
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153455-A2.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 22-DEC-2000; 2000WO-US35017.  
 XX  
 PR 23-DEC-1999; 99US-0471275.  
 PR 21-JAN-2000; 2000US-0486725.  
 PR 25-APR-2000; 2000US-0552317.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-457603/49.  
 DR N-PSDB; AAH99569.  
 XX  
 PT Isolated human polynucleotides encoding polypeptides, useful for the  
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 XX  
 XX Claim 20; Page 237; 1217pp; English.  
 XX  
 CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and

cells they are expressed in, such as: antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; viricide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; antiaggregat; haemostatic; vulnery; antitumor; osteopathic; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmune, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

XX Sequence 851 AA;

Query Match 100.0%; Score 1319; DB 22; Length 851;  
Best Local Similarity 100.0%; Pred No. 3.8e-98;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVGTDADGEGWPQVSLHALGOGHICGASLISPNNVLVSAAHCYIDDRGFRYSDPTQWTA 60  
Db 611 vvgtdadegewpwwsalhlggghicgalspnwlvsaahcyidrgfrysdptqwt 670  
Qy 61 FLGLHDSQSRAPQVQERRLKRISHPFNFDFDYDIALLELEKPAEYSSMVRPCLPD 120  
Db 671 flghdsqrsapvqerrlkrishpfnfdftdydiallelekpaeysmvrpiclpd 730  
Qy 121 ASHVFPAGKAIWVTGHTGYGGTGALILQKGEIRVINTTCENLLPQQITPRMCMVGFL 180  
Db 731 ashvfpagkaiwvtgthtgyggtgalilqkgeirvinttcenllpqqitprmmcvgl 790  
Qy 181 SGGVDSQCGSGGLSSVEADGRIFQAGVYVSWDGCAGQAKNKPQVYTRLPFRDWIKENTG 240  
Db 791 sggvdsqcgsgglssveadgrifqagvyswdgcagqaknkpvytrlpfrdwikentg 850  
Qy 241 V 241  
Db 851 V 851

# RESULT 6

AA06671  
ID AAY06671 standard; Protein; 855 AA.

XX AC AAY06671;

XX DT 09-NOV-1999 (first entry)

XX DE Tumour antigen derived gene-15 (TADG-15) protein.

XX KW Tumour antigen derived gene-15; TADG-15; serine protease; human;

XX KW breast cancer; ovary cancer; carcinoma; diagnosis.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Domain 1..54

FT /note= "cytoplasmic domain"

FT Domain 55..213

FT /note= "transmembrane domain"

FT Region 214..447

FT /note= "CUB repeat"

FT Region 453..602

FT /note= "ligand-binding repeat (class A motif)"

FT Domain 615..855

FT Modified-site 109..111 /note= "catalytic domain"

FT Modified-site 302..304 /note= "Asn is N-glycosylated"

FT Region 481..483 /note= "Asn is N-glycosylated"

FT Region 518..520 /note= "conserved SDE motif"

FT Region 554..556 /note= "conserved SDE motif"

FT Region 597..599 /note= "conserved SDE motif"

FT Cleavage-site 614..615 /note= "conserved SDE motif"

W09942120-A1.

26-AUG-1999.

18-FEB-1999; 99WO-US03436.

20-FEB-1998; 98US-0027337.

(UYAR-) UNIV ARKANSAS.

O'Brien TJ, Tanimoto H;

WPI; 1999-527418/44.

N-PSDB; AAX87815.

A new extracellular serine protease for diagnosis of neoplastic disease

Claim 3; Fig 10; 7lpp; English.

The present sequence represents a novel human extracellular serine protease, termed tumour antigen derived gene-15 protein (see AAY06671), that is overexpressed in breast and ovarian carcinomas. The TADG-15 gene (see AAX87815) can be used as a diagnostic and therapeutic target in ovarian carcinoma and other carcinomas including breast, prostate, lung and colon. The ligand binding domains of TADG-15 may be valuable in the uptake of specific molecules into tumour cells. The invention also provides: a vector that is capable of expressing DNA encoding TADG-15 protein; host cells selected from bacterial cells (especially *Escherichia coli*), mammalian cells, plant cells and insect cells; and a method of detecting expression of TADG-15 protein using a hybridisation probe.

XX Sequence 855 AA;

Query Match 100.0%; Score 1319; DB 20; Length 855;  
Best Local Similarity 100.0%; Pred No. 3.8e-98;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVGTDADGEGWPQVSLHALGOGHICGASLISPNNVLVSAAHCYIDDRGFRYSDPTQWTA 60

Db 615 vvgtdadegewpwwsalhlggghicgalspnwlvsaahcyidrgfrysdptqwt 674

Qy 61 FLGLHDSQSRAPQVQERRLKRISHPFNFDFDYDIALLELEKPAEYSSMVRPCLPD 120

Db 675 flghdsqrsapvqerrlkrishpfnfdftdydiallelekpaeysmvrpiclpd 734

Qy 121 ASHVFPAGKAIWVTGHTGYGGTGALILQKGEIRVINTTCENLLPQQITPRMCMVGFL 180

Db 735 ashvfpagkaiwvtgthtgyggtgalilqkgeirvinttcenllpqqitprmmcvgl 794

Qy 181 SGGVDSQCGSGGLSSVEADGRIFQAGVYVSWDGCAGQAKNKPQVYTRLPFRDWIKENTG 240

Db 795 sggvdsqcgsgglssveadgrifqagvyswdgcagqaknkpvytrlpfrdwikentg 854

Qy 241 V 241

Db 855 v 855

RESULT 7

AAB19552

ID AAB19552 standard; Protein; 855 AA.

XX AC AAB19552;

XX DT 22-JAN-2001 (first entry)

XX DE Human matriptase.

XX KW Matriptase; serine protease; human; breast cancer; pre-malignancy; actinic keratosis; leukoplakia; Barrett's epithelium; columnar metaplasia; ulcerative colitis; Bowenoid papulosis; adenomatous colorectal polyp; Oyerat erythroplasia; vulvar intraepithelial neoplasia; tumour; metastasis; therapy.

XX OS Homo sapiens.

XX FH Key

XX FT Region

XX FT /note= "complement subcomponent 1r and 1s (Clr/s)

XX FT /note= "region"

XX FT Region

XX FT /note= "complement subcomponent 1r and 1s (Clr/s)

XX FT /note= "region"

XX FT Region

XX FT /note= "LDL-receptor type region"

XX FT Region

XX FT /note= "LDL-receptor type region"

XX FT Region

XX FT /note= "LDL-receptor type region"

XX FT Domain

XX FT /label= Catalytic\_domain

XX FT Modified-site

XX FT /note= "N-glycosylated"

XX FT Modified-site

XX FT /note= "N-glycosylated"

XX FT Peptide

XX FT /note= "RGD motif"

XX FT Cleavage-site

XX FT /note= "conserved proteolytic activation site"

XX PN WO200053232-A1.

XX 14-SEP-2000.

XX 10-MAR-2000; 2000WC-US06111.

XX 12-MAR-1999; 99US-0124006.

XX (GEOU ) UNIV GEORGETOWN.

XX PI Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;

XX DR WPI: 2000-594268/56.

XX DR N-PSDB; AAB88493.

XX TT Treating malignancies, premalignant and pathologic conditions in a subject, comprises administering matriptase modulating agent -

XX PS Claim 14; Fig 16; 116pp; English.

XX CC The present sequence is that of the full-length form of human matriptase, a trypsin-like protease. This has an additional 172 amino acids compared with the truncated form of matriptase given in AAB19551. Either form can be produced in transfected or transfected cells using a claimed method. The zymogen (inactive)

CC form of matriptase is a single-chain protein. The active 2-chain form strongly interacts with fragments of a Kunitz-type serine protease inhibitor (hepatocyte growth factor activator inhibitor, HA-1) to form SDS-stable complexes. In breast cancer cells, matriptase is present mainly as the uncomplexed form. Only the complexed matriptase is detected in human milk. The invention is directed to a method of detecting a malignancy or a pre-malignant lesion in breast or other tissue by detecting the presence of single- or 2-chain forms of matriptase in the tissue. The object is to inhibit tumour onset, tumour growth and metastasis. CC Malignancies and pre-malignant conditions characterised by expression of the zymogen or activated form of matriptase are treated by administering an inhibitor of matriptase, especially a Bowman-Birk inhibitor. The pre-malignant condition is atypical ductal hyperplasia of the breast, actinic keratosis, leukoplakia, Barrett's epithelium of the esophagus, ulcerative colitis, adenomatous colorectal polyps, erythroplasia of the Queyrat, Bowen's disease, Bowenoid papulosis, vulvar intraepithelial neoplasia or dysplastic changes to the cervix. The invention also provides methods for in vivo or in vitro diagnosis of malignancy or pre-malignant lesion, and methods of identifying matriptase modulators, including activators and inhibitors.

XX SQ Sequence 855 AA;

Query\_Match 100.0%; Score 1319; DB 21; Length 855;

Best Local Similarity 100.0%; Pred. NO. 3.8e-98; Indels 0; Gaps 0;

Matches 241; Conservative 0; Mismatches 0;

QY 1 VVGTDADGEPMQVSLHALGQHICGASLSPNWLVSAAHYIDRGFRYSPTQMTA 60

Db 615 VVGTDADGEPMQVSLHALGQHICGASLSPNWLVSAAHYIDRGFRYSPTQMTA 674

QY 61 FLGLDQSQRSAPGVQERRLKRIISHPPFNDFDYDIALLELEKPAEYSSMVRPICLPD 120

Db 675 FLGLDQSQRSAPGVQERRLKRIISHPPFNDFDYDIALLELEKPAEYSSMVRPICLPD 734

QY 121 ASHVFPAGKAIWVTGWHTQYGGTGALILQKEIRVINQTTCENLFPQITPRMVCYGL 180

Db 735 ASHVFPAGKAIWVTGWHTQYGGTGALILQKEIRVINQTTCENLFPQITPRMVCYGL 794

QY 181 SGVSCQDSCGPLSSVEADGRIFQAGVSWGDCGCAQRNPGVYTRPLPRDWIKENTG 240

Db 795 SGVSCQDSCGPLSSVEADGRIFQAGVSWGDCGCAQRNPGVYTRPLPRDWIKENTG 854

QY 241 V 241

Db 855 V 855

RESULT 8

AAB06930

ID AAB06930 standard; Protein; 855 AA.

XX AC AAB06930;

XX DT 16-OCT-2001 (first entry)

XX DE Human membrane-type serine protease (MTSP) 1.

XX KW Human; transmembrane serine protease; membrane-type serine protease; MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic; lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy; matriptase.

XX OS Homo sapiens.

XX FH Key

XX FT Domain

XX FT /label= Protease\_domain

XX PN WO200157194-A2.

XX PD 09-AUG-2001.  
XX DE  
XX PF 02-FEB-2001; 2001WO-US03471.  
XX PR 03-FEB-2000; 2000US-0179982.  
XX PR 18-FEB-2000; 2000US-0183542.  
XX PR 22-JUN-2000; 2000US-0213124.  
XX PR 26-JUL-2000; 2000US-0220970.  
XX PR 08-SEP-2000; 2000US-0657986.  
XX PR 22-SEP-2000; 2000US-0234840.  
XX PA (CORV-) CORVAS INT INC.  
XX PI Madison EL, Ong BO, Yeh J;  
XX DR WPI; 2001-488877/53.  
XX DR N-PSDB; AAD13113.  
XX PT Novel single chain polypeptide comprising protease domain of type-II  
XX PT membrane-type serine protease or its catalytically active portion  
XX PT useful for treating and preventing cancer and tumor -  
XX PS Claim 12; Page 195-197; 256pp; English.  
XX CC The invention relates to transmembrane serine proteases and their  
XX CC corresponding nucleotides and the protease domain of a type-II  
XX CC membrane-type serine protease (MTSP). MTSP is useful for identifying  
XX CC compounds that modulate or inhibit its proteolytic activity and for  
XX CC formulating a medicament for treating neoplastic disease. MTSP and  
XX CC its corresponding nucleotides are useful in preventing or treating  
XX CC tumors or cancers such as lung carcinoma, colon adenocarcinoma and  
XX CC ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP  
XX CC is useful as a diagnostic marker for tumour development, growth and/or  
XX CC progression and as immunogens to generate antibodies that specifically  
XX CC bind to it. MTSP DNA is useful in a yeast two-hybrid system and in  
XX CC gene therapy. The present sequence is human MTSP1 protein (also called  
XX CC matriptase).  
XX SQ Sequence 855 AA;

Query Match 100.0%; Score 1319; DB 22; Length 855;  
Best Local Similarity 100.0%; Pred. No. 3.8e-98;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VVGGTDADEGENPWQVSLHALGQGHICGASLISPNWLYSAACHYIDDRGFYSPTQWTA 60  
DB 615 vvggtdadegewpqsllhalgqghicgasllspnlwlsaaahcyidrrfrysptqwt 674  
QY 61 FLGLHDQSORSAFGVQERLKRRIISHPFNFDFDYDIALLELEKPAEYSSMVRPCLPD 120  
DB 675 flglhdqsgsrsagvgqerrlkrliishpfndfddyiallelekpaeysmvrpiclpd 734  
QY 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTCENLLPQGITPRMCMVGFL 180  
DB 735 ashvfpagkaiwvtgwhgthqygggtgalllqkeirvinqtccenllpqgitprmcvfl 794  
QY 181 SGGVDSQCGDSGGLSSVEADGRIFQAGVVSWMGDCGCAQRNKPVGVTPLFRDWMKENTG 240  
DB 795 sggvdsqcgsggplssveadgrifqagvvsxwgdgcagrnkpgvytriplfrdwikentg 854  
QY 241 V 241  
DB 855 v 855  
RESULT 9  
AAB98500  
ID AAB98500 standard; Protein; 855 AA.  
XX AC AAB98500;  
XX XX

DT 03-AUG-2001 (first entry)  
XX DE Human TAGD-15.  
XX KW Human; TAGD-15; cytostatic; vaccine; ovarian tumour; cancer;  
XX KW tumour antigen-derived gene 15; extracellular serine protease.  
XX OS Homo sapiens.  
XX PN WO200129056-A1.  
XX XX 26-APR-2001.  
XX XX 20-OCT-2000; 2000WO-US29095.  
XX XX 20-OCT-1999; 99US-0421213.  
XX PA (UYAR-) UNIV ARKANSAS.  
XX PI O'Brien TJ, Tanimoto H;  
XX DR WPI; 2001-381031/40.  
XX DR N-PSDB; AAH23601.  
XX PT Novel extracellular serine protease, termed tumor antigen-derived gene  
XX PT 15 protein overexpressed in carcinomas and DNA encoding it, for  
XX PT diagnosis, treatment, prevention of cancer, particularly breast,  
XX PT ovarian cancer -  
XX PS Claim 11; Fig 2; 130pp; English.  
XX CC The present sequence represents human tumour antigen-derived gene 15  
XX CC (TAGD-15) protein. TAGD-15 is an extracellular serine protease. It was  
XX CC found that TAGD-15 is over-expressed in ovarian tumours. TAGD-15 protein  
XX CC or its fragments of 9-20 residues that lack TAGD-15 protease activity are  
XX CC useful for vaccinating an individual against TAGD-15, having, suspected  
XX CC of having or at risk of getting cancer. Furthermore, the TAGD-15 gene can  
XX CC be used as a diagnostic or therapeutic target in cancer.  
XX XX  
XX SQ Sequence 855 AA;

Query Match 100.0%; Score 1319; DB 22; Length 855;  
Best Local Similarity 100.0%; Pred. No. 3.8e-98;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VVGGTDADEGENPWQVSLHALGQGHICGASLISPNWLYSAACHYIDDRGFYSPTQWTA 60  
DB 615 vvggtdadegewpqsllhalgqghicgasllspnlwlsaaahcyidrrfrysptqwt 674  
QY 61 FLGLHDQSORSAFGVQERLKRRIISHPFNFDFDYDIALLELEKPAEYSSMVRPCLPD 120  
DB 675 flglhdqsgsrsagvgqerrlkrliishpfndfddyiallelekpaeysmvrpiclpd 734  
QY 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTCENLLPQGITPRMCMVGFL 180  
DB 735 ashvfpagkaiwvtgwhgthqygggtgalllqkeirvinqtccenllpqgitprmcvfl 794  
QY 181 SGGVDSQCGDSGGLSSVEADGRIFQAGVVSWMGDCGCAQRNKPVGVTPLFRDWMKENTG 240  
DB 795 sggvdsqcgsggplssveadgrifqagvvsxwgdgcagrnkpgvytriplfrdwikentg 854  
QY 241 V 241  
DB 855 v 855  
RESULT 10  
AAB35465  
ID AAB35465 standard; Protein; 855 AA.  
XX AC AAB35465;  
XX XX



DT 06-JUN-2001 (first entry)  
 XX Human membrane-type serine protease MT-SPL.  
 DE Human; membrane-type serine protease; MT-SPL; cancer.  
 XX Homo sapiens.  
 XX WO200123524-A2.  
 XX PD 05-APR-2001.  
 XX 02-OCT-2000; 2000WO-US27250.  
 XX 30-SEP-1999; 99US-0410362.  
 XX (REGC ) UNIV CALIFORNIA.  
 XX Craik CS, Takeuchi T, Shuman M;  
 XX WPI; 2001-245002/25.  
 XX N-PSDB; AAF28099.  
 XX New nucleic acid encoding a membrane type serine protease, useful for  
 PT the diagnosis, prognosis and treatment of cancer, particularly  
 PT metastatic cancers  
 XX Claim 7; Fig 1; 102pp; English.  
 XX The present invention provides the protein and coding sequences for the  
 CC novel human membrane-type serine protease MT-SPL. Increased expression of  
 CC this protein is associated with cancer, and so the sequences can be used  
 CC in cancer diagnosis and the identification of treatments. The present  
 CC sequence is the MT-SPL protein.  
 XX Sequence 855 AA;  
 SQ  
 Query Match 100.0%; Score 1319; DB 22; Length 855;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-98;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VVGTTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDRGRFYSDPTQWTA 60  
 DB 615 VVGTTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDRGRFYSDPTQWTA 674  
 QY 61 FLGLHDQSORGAPGVQERRKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPICLPD 120  
 DB 675 FLGLHDQSORGAPGVQERRKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPICLPD 734  
 QY 121 ASHVFPAGKAIWVTGWHTQYGGTGALLIQKEIRVINTTCENLLPQQITPRMCMVGFL 180  
 DB 735 ASHVFPAGKAIWVTGWHTQYGGTGALLIQKEIRVINTTCENLLPQQITPRMCMVGFL 794  
 QY 181 SGGVDSGCGDSSGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRPLFRDWIKENTG 240  
 DB 795 SGGVDSGCGDSSGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRPLFRDWIKENTG 854  
 QY 241 V 241  
 DB 855 v 855  
 RESULT 11  
 AAW22987  
 ID AAW22987 standard; Protein; 241 AA.  
 XX  
 AC AAW22987;  
 XX  
 DT 08-OCT-1997 (first entry)  
 XX Human serine protease 67 (SP67).  
 DE  
 XX

KW Human; colon carcinoma; COLO 201; cell line; serine protease; SP67;  
 XX screening; inhibitor; treatment; disease.  
 XX Homo sapiens.  
 XX JP09149790-A.  
 XX 10-JUN-1997.  
 XX 24-JUL-1996; 96JP-0212196.  
 XX 29-SEP-1995; 95JP-0275105.  
 XX (SUNR ) SUNTORY LTD.  
 XX WPI; 1997-357902/33.  
 XX N-PSDB; AAT79128.  
 XX Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67  
 PT - useful to screen for specific inhibitors, e.g. to search for, or  
 PT study agent for treatment of various diseases  
 XX Claim 1; Pages 12-13; 16pp; Japanese.  
 XX The present sequence is the human colon carcinoma COLO 201  
 CC cell line derived serine protease 67 (SP67), which can be used to  
 CC screen for specific inhibitors, e.g. to search for, or study an  
 CC agent for the treatment of various diseases.  
 XX Sequence 241 AA;  
 SQ  
 Query Match 99.7%; Score 1315; DB 18; Length 241;  
 Best Local Similarity 99.6%; Pred. No. 2.3e-98;  
 Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VVGTTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDRGRFYSDPTQWTA 60  
 DB 1 VVGTTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDRGRFYSDPTQWTV 60  
 QY 61 FLGLHDQSORGAPGVQERRKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPICLPD 120  
 DB 61 FLGLHDQSORGAPGVQERRKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPICLPD 120  
 QY 121 ASHVFPAGKAIWVTGWHTQYGGTGALLIQKEIRVINTTCENLLPQQITPRMCMVGFL 180  
 DB 121 ASHVFPAGKAIWVTGWHTQYGGTGALLIQKEIRVINTTCENLLPQQITPRMCMVGFL 180  
 QY 181 SGGVDSGCGDSSGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRPLFRDWIKENTG 240  
 DB 181 SGGVDSGCGDSSGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRPLFRDWIKENTG 240  
 QY 241 V 241  
 DB 241 v 241  
 RESULT 12  
 ABG21442  
 ID ABG21442 standard; Protein; 932 AA.  
 XX  
 AC ABG21442;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #21433.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 XX

XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YF;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS85629.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 20; SEQ ID No 51801; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG0010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 932 AA;

Query Match 93.0%; Score 1227; DB 22; Length 932;  
Best Local Similarity 93.8%; Pred. No. 1.1e-90;  
Matches 228; Conservative 4; Mismatches 9; Indels 2; Gaps 2;  
Qy 1 VVGSTDADEGEWPQVSLHALGOGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWT- 59  
Db 590 vvgstddadegewpqlslhalgghicgalslspnwlvsaahcfidrrfmsnptqwr 749  
Qy 60 AFLGLHDSQSRAP-GVQERRLKRIISHPFNFDFDYDIALLELEKPAEYSSMVRPCL 118  
Db 750 aflglhdqsgrrppwgcrrlkrliishpfndftfdydiallelekpaeysmvrpcl 809  
Qy 115 PDASHVPAGKAIWVTGWGHTQYGGTGALILQKEIRVINTTCENLLPQQTIPRMKCVG 178  
Db 810 pdashvpagkaiwvtgwghtqyggtgallilqkeirvinttcenllpqqtprmmcvg 869  
Qy 179 FLGGVDSQSGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKEN 238  
Db 870 flsggvdsqsgdggplssveadgrifqagvswsvegcvgrnkpgvytrlpfrdwiken 929  
Qy 239 TGV 241  
Db 930 tgv 932  
RESULT 13

AAB98507  
ID AAB98507 standard; Protein; 902 AA.  
XX AC AAB98507;  
XX DT 03-AUG-2001 (first entry)  
XX DE Murine epithin.  
XX KW Murine; TAGD-15; cytostatic; vaccine; ovarian tumour; cancer; epithin;  
XX KW tumour antigen-derived gene 15; serine protease.  
XX OS Mus musculus.  
XX PN WO200129056-A1.  
XX PD 26-APR-2001.  
XX PF 20-OCT-2000; 2000WO-US29095.  
XX PR 20-OCT-1999; 99US-0421213.  
XX PA (UYAR-) UNIV ARKANSAS.  
XX PI O'Brien TJ, Tanimoto H;  
XX DR WPI; 2001-381031/40.  
XX PT Novel extracellular serine protease, termed tumor antigen-derived gene  
XX PT 15 protein overexpressed in carcinomas and DNA encoding it, for  
XX PT diagnosis, treatment, prevention of cancer, particularly breast,  
XX PT ovarian cancer -  
XX PS Disclosure; Page 97-99; 130pp; English.  
XX CC The present invention relates to human tumor antigen-derived gene 15  
XX CC (TAGD-15) protein and coding sequence (see AAB98507 and AAB98500).  
XX CC TAGD-15 is an extracellular serine protease. It was found that TAGD-15 is  
XX CC over-expressed in ovarian tumours. TAGD-15 protein or its fragments of  
XX CC 9-20 residues that lack TAGD-15 protease activity are useful for  
XX CC vaccinating an individual against TAGD-15, having, suspected of having or  
XX CC at risk of getting cancer. Furthermore, the TAGD-15 gene can be used as a  
XX CC diagnostic or therapeutic target in cancer. The present sequence was used  
XX CC in a sequence homology alignment with TAGD-15.  
XX SQ Sequence 902 AA;  
Query Match 83.4%; Score 1100; DB 22; Length 902;  
Best Local Similarity 86.9%; Pred. No. 1.8e-80;  
Matches 199; Conservative 17; Mismatches 13; Indels 0; Gaps 0;  
Qy 1 VVGSTDADEGEWPQVSLHALGOGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWT 60  
Db 615 vvgstddadegewpqlslhalgghicgalslspnwlvsaahcfidrrfmsnptqwr 674  
Qy 61 FLGLHDSQSRAP-GVQERRLKRIISHPFNFDFDYDIALLELEKPAEYSSMVRPCL 120  
Db 675 flglhdqskrsagvqcelklrliishpfndftfdydialleleksveystvvrpiclpd 734  
Qy 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINTTCENLLPQQTIPRMKCVGL 180  
Db 735 athvfpagkaiwvtgwghtqyggtgallilqkeirvinttcenllpqqtprmmcvgl 794  
Qy 181 SGGVDSQSGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLP 229  
Db 795 sggvdsqsgdggplssveadgrifqagvswsvegcvgrnkpgvytrlp 843  
RESULT 14  
ABG21439  
ID ABG21439 standard; Protein; 129 AA.  
XX

AC ABG21439;  
XX  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #21430.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR  
XX 23-AUG-2000; 2000US-0649167.  
PR  
XX (HYSE-) HYSSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX WPI: 2001-639362/73.  
DR  
XX N-PSDB; AAS85626.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PT  
PS Claim 20; SEQ ID NO 51798; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (II) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 129 AA;  
SQ  
Query Match 48.7%; Score 643; DB 22; Length 123;  
Best Local Similarity 92.2%; Pred. No. 1.6e-44;  
Matches 119; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
QY 106 PAYSSWVRPCLPASHVFPAGKAIWVTGHTQYGGTGALIIQKEIRVINTTCENL 165  
DB 1 payesswvrpclpashvfpagkaiwvtgwtgqyggtagllqkgeirvinttce 60  
QY 166 LPQQTTPRMVCGFFSGGVDSQSGPLSVBAGDRIFQAGVYVWGSDGCAORNKPGVY 225  
DB 61 lpqqltpmrmvcmvffsggvdsdsssssspsvveadgrifagvrvsgdgcagrnkpgv 120  
QY 226 TRLPFRDW 234  
|||||:|

Db 121 tripfqdw 129  
RESULT 15  
AAE01942  
ID AAE01942 standard; Protein; 233 AA.  
XX  
AC AAE01942;  
XX  
DT 31-JUL-2001 (first entry)  
XX  
DE Human endotheliase 1 protease domain.  
XX  
KW Human; endotheliase 1; protease domain; cytostatic; vulnerary; wound;  
KW neotropic; periodontitis; dermatological disorder; gene therapy; scar;  
KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;  
KW chronic inflammatory disease; ocular disorder; circulatory disorder;  
KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;  
KW liver cirrhosis, osteoradionecrosis, systemic sclerosis; oesophageal;  
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;  
KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn.  
XX  
OS Homo sapiens.  
XX  
XX WO200136604-A2.  
PN  
XX 25-MAY-2001.  
PD  
XX 17-NOV-2000; 2000WO-US31803.  
XX  
XX 18-NOV-1999; 99US-0166391.  
PF  
XX 22-SEP-2000; 2000US-0234840.  
PR  
XX (CORV-) CORVAS INT INC.  
PA  
XX Madison EL, Ong EO;  
PI  
XX WPI: 2001-336001/35.  
DR  
XX N-PSDB; AAD05795.  
DR  
XX New nucleic acid encoding a protein comprising endotheliase activity  
PT useful in the prevention and treatment of e.g. vascular malformations,  
PT cardiovascular disorders, and chronic inflammatory disease -  
XX  
XX Claim 4; Page 134-135; 152pp; English.  
XX  
XX The present sequence is human endotheliase 1 protease domain.  
CC The invention relates to an endotheliase protein, endotheliase protease  
CC domain and their corresponding nucleic acid molecules. An endotheliase  
CC protein or protease domain of it is useful for the treatment and  
CC diagnosis of disorders associated with aberrant angiogenesis or undesired  
CC neovascularisation. The undesired angiogenesis is associated with  
CC disorders selected from solid neoplasm, vascular malformations and  
CC cardiovascular disorders such as angiofibroma, angiolipoma,  
CC atherosclerosis, restenosis/reperfusion injury, arteriovenous  
CC malformations, haemangiomas and vascular adhesions, dyschondroplasia  
CC with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic  
CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau  
CC syndrome, chronic inflammatory diseases such as diabetes mellitus,  
CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,  
CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,  
CC granulations/burns, hypertrophic scars, liver cirrhosis,  
CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and  
CC systemic sclerosis and aberrant wound repairs, circulatory disorders  
CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,  
CC dysmotility, sclerodactyly and telegiectasis, dermatological disorders  
CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,  
CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine  
CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome  
CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness  
CC caused by ocular neovascular disease, corneal graft neovascularisation,  
CC macular degeneration, retinopathy of prematurity, retrolental  
CC fibroplasia and corneal neovascularisation. The nucleic acids of the



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:08:24 ; Search time 19.03 Seconds  
(without alignments)  
1216.896 Million cell updates/sec

Title: US-09-657-986b-2

Perfect score: 1319

Sequence: 1 VVGTDADGEWPQVSLHA.....PGVYTRPLFRDWIKENTGV 241

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*

1: Pir1:\*

2: Pir2:\*

3: Pir3:\*

4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1151	87.3	855	JC7731	membrane-bound arg
2	1151	87.3	855	JC7775	membrane type-seri
3	552.5	41.9	1034	A53663	enteropeptidase (E
4	547.5	41.5	1019	A56318	enteropeptidase (E
5	544.5	41.3	1035	A43090	enteropeptidase (E
6	483	36.6	786	A47547	serine proteinase
7	477	36.2	638	KOMSP	plasma kallikrein
8	473.5	35.9	417	S0845	hepsin (EC 3.4.21.
9	472	35.8	790	PLPG	plasmin (EC 3.4.21
10	469	35.6	638	KORTPL	plasma kallikrein
11	467	35.4	275	S40007	trypsin (EC 3.4.21
12	466	35.3	416	S33777	hepsin (EC 3.4.21.
13	463.5	35.1	1524	T30337	polyprotein - Afri
14	462	35.0	275	S40005	trypsin (EC 3.4.21
15	462	35.0	638	KOHUP	plasma kallikrein
16	461	35.0	343	A57014	proctasin (EC 3.4.
17	457	34.6	274	S35339	trypsin (EC 3.4.21
18	456.5	34.6	267	S40006	trypsin (EC 3.4.21
19	455	34.5	625	1 KFHU1	coagulation factor
20	453.5	34.4	812	PLMS	plasmin (EC 3.4.21
21	452	34.3	460	B61545	plasmin (EC 3.4.21
22	452	34.3	810	1 PLHU	plasmin (EC 3.4.21
23	448.5	34.0	810	2 146260	plasmin (EC 3.4.21
24	446	33.8	238	1 TRWV5Y	trypsin-like prote
25	444	33.7	455	2 A61545	plasmin (EC 3.4.21
26	441	33.4	810	2 B30848	plasmin (EC 3.4.21
27	439.5	33.3	812	1 PUBO	plasmin (EC 3.4.21
28	439	33.3	270	2 S56160	mast cell tryptase
29	436.5	33.1	416	1 KFB0	coagulation factor

30 436 33.1 254 1 TRWV3Y trypsin-like prote  
31 436 33.1 1420 2 A32869 apolipoprotein(a)  
32 435.5 33.0 247 1 TRDG trypsin (EC 3.4.21  
33 434.5 32.9 4548 1 S00657 apoprotein(a) (EC  
34 431 32.7 273 2 A47246 tryptase (EC 3.4.2  
35 430 32.6 263 2 A21195 chymotrypsin (EC 3  
36 429.5 32.6 231 1 TRPGTR trypsin (EC 3.4.21  
37 429.5 32.6 461 1 KFHU coagulation factor  
38 429 32.5 242 2 S49489 trypsin (EC 3.4.21  
39 429 32.5 276 2 A38654 mast cell proteina  
40 428 32.4 277 2 S35340 trypsin (EC 3.4.21  
41 426.5 32.3 452 1 A30351 coagulation factor  
42 426 32.3 1113 2 JB0315 low-density lipopr  
43 425.5 32.3 247 1 B25852 trypsin (EC 3.4.21  
44 425.5 32.3 247 2 A27547 trypsin (EC 3.4.21  
45 425 32.2 275 2 A32410 tryptase (EC 3.4.2

#### ALIGNMENTS

RESULT 1

JC7731

membrane-bound arginine-specific serine proteinase precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 14-Dec-2001

C:Accession: JC7731

R:Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athaud

J. Biochem. 130, 425-430, 2001

A:Title: Characterization of a membrane-bound arginine-specific serine protease from

A:Reference number: JC7731; MUID:21421307; PMID:11530019

A:Accession: JC7731

A:Molecule type: mRNA

A:Residues: 1-855 <KIS>

A:Cross-references: DDBJ:AB049189

A:Experimental source: strain Male, 7-week-old

C:Comment: This enzyme, localized mainly on brushborder membranes of the intestine, p

C:Keywords: protein digestion

Query Match 87.3%; Score 1151; DB 2: Length 855;  
Best Local Similarity 87.1%; Pred. No. 7,4e-99;  
Matches 210; Conservative 15; Mismatches 16; Indels 0; Gaps 0;

QY 1 VVGTDADGEWPQVSLHALCGGHICGASLISPNWLSAAHCYIDDRGFRYSDEPTQWTA 60

Db 615 VVGTDADGEWPQVSLHALCGGHICGASLISPNWLSAAHCYIDDRGFRYSDEPTQWTA 674

QY 61 FLGLHDQSORSPAGVQERLKRITSHPPFNDFDYDIALLEKPAEYSSMVRPCLPDP 120

Db 675 FLGLHDQSORSPAGVQERLKRITSHPPFNDFDYDIALLEKPAEYSSMVRPCLPDP 734

QY 121 ASHVFPAGKAIWVTGMGHTQYGGTGALILQKGEIRVINOTTCEILLPQIIPRMVCVGL 180

Db 735 NTHVFPAGKAIWVTGMGHTQYGGTGALILQKGEIRVINOTTCEILLPQIIPRMVCVGL 794

QY 181 SGGVDSGDSGGLSSVEADGRIFQAGVSWGDCGACORNPQVYTRPLPFRDWIKENTG 240

Db 795 SGGVDSGDSGGLSSVEADGRIFQAGVSWGDCGACORNPQVYTRPLPFRDWIKENTG 854

QY 241 V 241

Db 855 V 855

RESULT 2

JC7775

membrane type-serine protease 1 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: JC7775

R:Satomi, S.; Yamazaki, Y.; Tsuzuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.

Biochem. Biophys. Res. Commun. 287, 995-1002, 2001



F:785-1019/Domain: enteropeptidase light chain #status predicted <LCH>  
F:785-1014/Domain: trypsin homology <TRY>  
F:116-147/179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:  
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted  
F:825-876,971/Active site: His, Asp, Ser #status predicted

Query Match 41.5%; Score 547.5; DB 1; Length 1019;  
Best Local Similarity 44.7%; Pred. No. 1.2e-42;  
Matches 106; Conservative 44; Mismatches 80; Indels 7; Gaps 4;

QY 1 VVGTTDADGEPMQVSLHALCGGHICGASLISPNMLVSAACHYIDDRGRFYSPTQWTA 60  
Db 785 IVGGSNAKEGAPWVGLY-YGGRLLCGASLVSSDLVLSAAHCYVYV-----RNLEPSKMTA 839  
QY 61 FLGLHQDSQRSAPGVQERRLKRIISHPPFNDFDYDIALLEKPAEYSSMVRPCLPD 120  
Db 840 ILGLHKNLTSPTQVPRIDEIVNPHNRRKNDLMMHLEKPVNYTYDIQICLPE 899  
QY 121 ASHVFPAGKAIWVGHTQYCGTGALILQKGEIRVINOTTENLLPO-QITPRMCMVGF 179  
Db 900 ENQVFPGRNCSTAGTGVYVGGTANILQEADVPLLSNERCQQQOMPEYNTENNICAGY 959  
QY 180 LSGGVSCDGGSGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYTRPLFRDWIK 236  
Db 960 EEGIDSCDGGSGPLMCOE-NNRMLAGVTSGYKCALPNRPGVYARVSRFTWIO 1015

RESULT 5  
A43090  
N:enteropeptidase (EC 3.4.21.9) precursor - bovine  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A43090; A48874; A61436  
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.  
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994  
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic composed of two distinct genes  
A:Reference number: A43090; MUID:94329561  
A:Accession: A43090  
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1035 <KIT>  
A:Cross-references: GB:U09859; NID:g746410; PIDN:AAB40026.1; PID:g746411  
A:Experimental source: small intestine  
R:LaValle, E.R.; Rehemtulla, A.; Racine, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.; J. Biol. Chem. 268, 23311-23317, 1993  
A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of human enterokinase  
A:Reference number: A48874; MUID:94043122  
A:Accession: A48874  
A:Molecule type: protein  
A:Residues: 801-1035 <LAV>  
A:Cross-references: GB:L19663; NID:g416131; PIDN:AAA16035.1; PID:g416132  
A:Note: parts of this sequence, including the amino end of the mature protein, were confirmed by N-terminal sequencing  
R:Light, A.; Janska, H. J. Protein Chem. 10, 475-480, 1991  
A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.  
A:Reference number: A61436; MUID:92189715  
A:Accession: A61436  
A:Molecule type: protein  
A:Residues: 801-807, 'y', 809-827 <LIG>  
C:Comment: The mechanism of association with the membrane of the intestinal brush border membrane attachment using a signal-anchor sequence.  
C:Comment: Conversion from membrane-bound to soluble forms may involve further processing  
C:Complex: mature enteropeptidase is variously reported to contain two (heavy and light) lflde linked  
C:Function: cleaves propeptide from trypsinogen to produce active trypsin  
A:Description: intestinal digestive hydrolase cascade  
C:Pathway: enteropeptidase; Ctr/Clis repeat homology; LDL receptor ligand-binding repeat  
C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein  
F:22-38/Domain: transmembrane #status predicted <TM>  
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>  
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL>  
F:358-520/Domain: MAM homology <MAM>  
F:542-647/Domain: Ctr/Clis repeat homology <CTR>  
F:559-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F:594-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SC>  
F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>  
F:801-1030/Domain: trypsin homology <TRY>  
F:116-147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/BLN  
F:788-912,826-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted  
F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 41.3%; Score 544.5; DB 1; Length 1035;  
Best Local Similarity 43.5%; Pred. No. 2.3e-42;  
Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 1 VVGTTDADGEPMQVSLHALCGGHICGASLISPNMLVSAACHYIDDRGRFYSPTQWTA 60  
Db 801 IVGGSNREGAPWVVALYFDDQ-QVCGASLVSRDLVLSAAHCYVYV---RNMEPSKWA 855  
QY 61 FLGLHQDSQRSAPGVQERRLKRIISHPPFNDFDYDIALLEKPAEYSSMVRPCLPD 120  
Db 856 VLGLHMASNLTSPTQVPRIDEIVNPHNRRKNDLMMHLEKPVNYTYDIQICLPE 915  
QY 121 ASHVFPAGKAIWVGHTQYCGTGALILQKGEIRVINOTTENLLPO-QITPRMCMVGF 179  
Db 916 ENQVFPGRNCSTAGTGVYVGGTANILQEADVPLLSNERCQQQOMPEYNTENNICAGY 975  
QY 180 LSGGVSCDGGSGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYTRPLFRDWIK 236  
Db 976 EAGVDSCDGGSGPLMCOE-NNRMLAGVTSGYKCALPNRPGVYARVSRFTWIO 1031

RESULT 6  
A47547  
C:Species: Drosophila melanogaster  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A47547  
R:Appel, L.F.; Prout, M.; Abu-Shumays, R.; Hammonds, A.; Garbe, J.C.; Fristrom, D.; F. Proc. Natl. Acad. Sci. U.S.A. 90, 4937-4941, 1993  
A:Title: The Drosophila Stubble-stubblid gene encodes an apparent transmembrane serine proteinase  
A:Reference number: A47547; MUID:93281671  
A:Accession: A47547  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-786 <APP>  
A:Cross-references: GB:L11451; NID:g158511; PIDN:AAA28918.1; PID:g158512  
C:Genetics:  
A:Gene: Sb-sbd  
A:Cross-references: FlyBase:FBgn000319  
C:Superfamily: serine proteinase stubble-stubblid; trypsin homology  
C:Keywords: hydrolase; serine proteinase; transmembrane protein  
F:61-77/Domain: transmembrane #status predicted <TM>  
F:543-781/Domain: trypsin homology <TRY>

Query Match 36.6%; Score 483; DB 1; Length 786;  
Best Local Similarity 39.2%; Pred. No. 8.9e-37;  
Matches 98; Conservative 47; Mismatches 85; Indels 20; Gaps 8;

QY 1 VVGTTDADGEPMQVSLH-----ALGQGHICGASLISPNMLVSAACHYIDDRGRFYSDP 55  
Db 543 IVGGSNAKEGAPWVVALYFDDQ-QVCGASLVSRDLVLSAAHCYVYV---RNMEPSKWA 855  
QY 56 TQWTAFLGHDS--ORSAPGVQERRLKRIISHPPFNDFDYDIALLEKPAEYSSM 113  
Db 597 SQIRIVGVDISHVQEQPYVGVVAKV--HPKISFLTYEIDIALVKLEPLEFAPHV 655  
QY 114 RPICLPDASHVFPAGKAIWVGHTQYCGTGALILQKGEIRVINOTTENLL-----PQQ 169  
Db 656 SPICLPETDSSL-IGNNATVTCWGRSLSEGGLTSLVQLQEVSPVSNDCNCKSMENRAGQE 714

```

QY 170 ITPRM-MCVGFLSGVDSCQDGGPLSSVEAGRTFQAGVSWGDCGCAORNPGVVTRL 228
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 715 FIPDFLCAGYETGGQDSCQDGGPLQAKSQDGRFFLAGIISWGIGCAEANLPGVCTRI 774
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 229 PLFRDWIKEN 238
      | : | | :
Db 775 SKTFPWILEH 784
      | : | | :

RESULT 7
KQMSPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: A36557
R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemon
DNA Cell Biol. 9, 737-748, 1990
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison
A:Reference number: A36557; MUID:91090844
A:Accession: A36557
A:Molecule type: mRNA
A:Residues: 1-638 <SEI>
A:Cross-references: GB:M58588; NTD:G200358; PIDN:AAA63393.1; PID:G200359
A:Note: part of this sequence, including the amino ends of both the heavy and light chain
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C:Comment: The zymogen is activated by factor XIa, which cleaves the molecule into a li
are linked by one or more disulfide bonds.
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-300/Product: plasma kallikrein heavy chain #status experimental <HC>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <LC>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-111,194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
F:127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

```

Query Match 36.2%; Score 477; DB 1; Length 638;  
Best Local Similarity 39.3%; Pred. NO. 2.5e-36;  
Matches 96; Conservative 43; Mismatches 91; Indels 14; Gaps 7;

Qy	1	VYGGTDAGEWPQVSIHA--LGQHICGASLISPNMLVSAAHCYIDDRFRYSDDPTQM	58
		:   :       :  :   :  :   :  :   :	
Db	391	IYVGTFNASLGEWPMQVSLQVKLSQTHLCGSGIIGROWLTAACHCF---DGIPY--PDVV	445
Qy	59	TAFGLGHDSO--RSAPGVQRKLIIHSHPFNDFTFDYDIALLEKPAEYSSMVRI	116
		: :  : :  : :  : :  : :  : :  : :  : :  : :	
Db	445	RIYGILSLSIETKP---SSRIKELIHQBKYKVSBNYDIALIKLTPLNTEFQKI	502
Qy	117	CLPDAASHVPAGAIWTGWHTGYGGTGALIQKEIRVINQTTCENLLPQQ-IITPRM	175
		:::   :    :   :    :   :    :   :    :   :	
Db	503	CLPSKADNTIYTNCWTTGWGTKEOGTQLQATPLVPNECCKKYRDVINKQMI	562
Qy	176	CVGFLLSGVDSCODSGSPSSYEADGRIFQAQVYGSWDGCQAORNKPGVVTLPFLRDM	235
		:   :    :    :    :    :    :    :    :    :    :	
Db	563	CAGYEGGTDACKDSGGPL-VCKHSGRWQLVGITSWGCGCRKDQPEVITKVSEYNMWI	621
Qy	236	KENT	239
Db	622	LKRT	625

RESULT 8  
S00845  
hepsin (EC 3.4.21.-) - human  
C:Species: Homo sapiens (man)  
C:date: 31-Dec-1988 #sequence revision 31-Dec-1988 #text change 18-Jun-1999

C:Accession: S00845  
 R:Leytus, S.P.; Lobb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.  
 Biochemistry 27, 1067-1074, 1988  
 A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane  
 A:Reference number: S00845; MUID:88209431  
 A:Accession: S00845  
 A:Molecule type: mRNA  
 A:Residues: 1-417 <LEV>  
 A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064  
 C:Genetics:  
 A:Gene: GDB:HPN; TMPRS1; hepsin  
 A:Cross-references: GDB:I35685; OMIM:142440  
 A:Map position: 19q11-19q13.2  
 C:Superfamily: hepsin; trypsin homology  
 C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein  
 F:23-45/Domain: transmembrane #status predicted <TMN>  
 F:163-400/Domain: trypsin homology <TRY>  
 F:188-204, 291-359,322-338,349-381/disulfide bonds: #status predicted  
 F:203,257,355/Active site:His, Asp, Ser #status predicted

```

Query Match      35.98; Score 473.5; DB 1; Length 417;
Best Local Similarity 40.7%; Pred. N.3.2e-36;
Matches 100; Conservative 38; Mismatches 89; Indels 19; Gaps

QY   1  VVGVDADGEWPQVSLHALGQCHICGASLISPNWLYSAAHCYIDDRGFYSDPDTOWTA 60
      :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db   163 IYGGEDTSLGRWPQVSLRYDG -AHLGGSGLLSGDWLTAAHCFE---RNRVLSPWRV 217

QY   61 FGLGHQSORGAPGVQERUKRIISH----PFF--NDTFDYDIALALELEKPAEYSMWR 114
      :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db   218 FAGA--VAQASPGLQ-LGVQAVYHGVLPPFRDNSENSNDIALVHLSPLFTEYIQ 274

QY   115 PCLPDASHVFPAKAIWTGWGTGYGTGALILQKEIRVINOTTC E--NLLPOOITP 172
      :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db   275 PVCLPAAGCALVDGKICTVTGWNTQYGCQAGVLQEARVPISIINDVCNGADFVGNQIKP 334

QY   173 RMCMVGFLSGGVDSOGDSGGPL---SSVEADGRIFQAGVYSWGDCGAQRNKGVTRLP 229
      :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db   335 KMFCAYPEGGIDACQDSGGPFVCEDETSIRTPRWRLCGIVSWGTCGACALACKPEVTKVS 394

QY   230 LFRDWII 235
      :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db   395 DFEWI 400

```

## RESULT

```

RESULT 9
PLPG
  plasmin (EC 3.4.21.7) precursor - pig (fragment)
N:Alternate names: plasminogen
N:Contains: miniplasminogen
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 07-Sep-1990 #sequence_revision 01-Nov-1996 #text_change 18-Jul-1997
C:Accession: S03733; S03737; A25834
R:Schaller, J.; Marti, T.; Rosselet, S.J.; Kaempfer, U.; Rickli, E.E.
  Fibrinolysis 1, 91-102, 1987
A:Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the
  A:Reference number: S03733
A:Accession: S03733
A:Molecule type: protein
A:Residues: 1-560 <SCH>
R:Brunscholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg,
  Eur. J. Biochem. 114, 465-470, 1981
A:Title: Comparison of the primary structure of the N-terminal CNBr fragments of huma
  A:Reference number: S03735; MUID:81212097
A:Accession: S03737
A:Molecule type: protein
A:Residues: 1-57 <BRU>
R:Marti, T.; Schaller, J.; Rickli, E.E.
  Eur. J. Biochem. 145, 279-285, 1985
A:Title: Determination of the complete amino-acid sequence of porcine miniplasminogen
  A:Reference number: A25834; MUID:85203907
A:Accession: A25834

```



A:Molecule type: protein  
A:Residues: 450-790 <MAR>  
C:Function:  
N:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a variety of tissues; also activates the urokinase-type plasminogen activator system.  
A:Pathway: fibrinolysis  
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasmin; serine protease  
F:1-790/Product: plasminogen #status predicted <PRO>  
F:1-77/Domain: plasminogen-related protein precursor homology (fragment) <PLPH>  
F:1-77/Domain: activation peptide #status predicted <APT>  
F:78-560/Product: plasmin chain A #status predicted <ACH>  
F:84-162/Domain: kringle homology <KRL>  
F:166-243/Domain: kringle homology <KR2>  
F:256-333/Domain: kringle homology <KR3>  
F:358-435/Domain: kringle homology <KR4>  
F:450-790/Product: miniplasminogen #status experimental <MTN>  
F:461-540/Domain: kringle homology <KR5>  
F:561-790/Product: plasmin chain B #status experimental <BCH>  
F:561-783/Domain: trypsin homology <TRY>  
F:30-54,34-42,84-162,105-145,133-157,166-243,169-297,187-226,215-238,256-333,277-316,305 bonds: #status predicted  
F:502,645,740/Active site: His, Asp, Ser #status predicted

Query Match 35.8%; Score 472; DB 1: Length 790;  
Best Local Similarity 41.8%; Pred No. 9,4e+36;  
Matches 100; Conservative 35; Mismatches 88; Indels 16; Gaps 6;

QY 1 VVGGTDAEGEPWQVSLHALGOGHICGLASLISPNMLVSAAHCYIDDRGFYSDPQTQWA 60  
||| : ||||| : ||| : ||||| : ||| : ||| : ||| :  
Db 561 VVGCVSIPHSWPQWSLRVYRGHFCCGLSPFWVLTAKHCL----EKSSPSISKY 615  
  
QY 61 FLGLHQSQRSAPGVORRLKIIRIHPFFDFDYLLELEKEPAEYSSMVRPCLPD 120  
||| : ||||| : ||| : ||||| : ||| : ||| : ||| :  
Db 616 ILCAHE-YILGGVGVEIDVSKFKPE-----SEADIALKLSSPAVITDKVIPACLP 668  
  
QY 121 ASHVFPAGKAIWYTGNGHTGYGTGALILQGBIRVINOTCE--NLPPQOITFRMCVG 178  
:|| : ||||| : ||| : ||||| : ||| : ||| : ||| :  
Db 669 PNVVADRATACYITGWGETK-GTYGALLKEARLPVIENKVCNKRYELGKGVPNELCAG 727  
  
QY 179 FLSGGVDSOCGDGGLSSLVSVEAGRIFQGVSWGDCGAQRNKFPGYVTRLPEDWIKE 237  
:||||| : ||||| : ||| : ||||| : ||| : ||| : ||| :  
Db 728 HLAGGIDSOCGDGGGLPVCEKDRIQLQ-GVTSWGGLCALPNKPGRVYRVSRVETWIEE 785

RESULT 10  
KORTPL  
plasma kallikrein (EC 3.4.21.34) precursor - rat  
N:Alternate names: Fleischer factor; kininogenin; serum kallikrein  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 30-Sep-1992 #sequence,revision 30-Sep-1992 #text\_change 18-Jun-1999  
R:Accession: A33180; A33320; S06851; I53041; S06852  
R:Beaubien, G.; Rosinski-Chupin, I.; Mattel, M.G.; Mbikay, M.; Chretien, M.; Seidah, N.G.  
Biochemistry 30, 1628-1635, 1991  
A:title: Gene structure and chromosomal localization of plasma kallikrein.  
A:Reference number: A39180; MUID:91129236  
A:Accession: A39180  
A:Molecule type: DNA  
A:Residues: 1-638 <BEA>  
A:Cross-references: GB:J05315  
A>Note: the authors translated the codon GAG for residue 81 as Glu  
R:Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazure,  
DNA 8, 563-574, 1989  
A:title: The cDNA structure of rat plasma kallikrein.  
A:Reference number: A33320; MUID:90091743  
A:Accession: A33320  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-638 <SEI>  
A:Cross-references: GB:M30282; NID:q205010; PIDN:AAA41463.1; PID:q205011  
A>Note: part of this sequence, including the amino ends of both the heavy and light chains.  
R:Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.

```
A:Residues: 1-275 <MUE>
A:Cross-references: EMBL:Z22930; NID:g410323; PIDN:CAA80515.1; PID:g410328
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase
F:49-269/Domain: trypsin homology <TRY>
F:89,134,230/Active site: His, Asp, Ser #status predicted

Query Match 35.4%; Score 467; DB 2; Length 275;
Best Local Similarity 39.7%; Pred. No. 8e-36;
Matches 96; Conservative 39; Mismatches 89; Indels 18; Gaps 6;

Qy 1 VVGTTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYIDDRGFRYSDPTQWTA 60
Db 49 IVGGEDVSTPTQVSLQVFN-NSHRCGGSVLAKWLTAAHCTVN-----LQPSLAV 101
Qy 61 FLGLHDQSQRSAPGVQERRLKRIISHFFNDFDIDIALLELEKPAEYSSMVRPICLPD 120
Db 102 RLG---SSRHASGGTVVRVARVLEHFNDDSTIDYDFSLMELETSLTSDVQVPSLPE 157
Qy 121 ASHVFPAGKATWVTGWHTQVGGTGALLQKGEIRVINQTTCEMLPQO--ITPRMCMVG 178
Db 138 QDEAVEDGTMTVTGWNTOGTAESAENALRAANIPTVNQKECTIAYSSGGITDRMLCAG 217
Qy 179 FLISGVDSCQDSGGPLSSVEADGRIFQAGVSWGDCQARNKPGVYTRPLFRDWIKEN 238
Db 218 YKRGKDACQDSGGPL--VVDGKL--VGWSWGTGFCAMPGYPGYVARVAVRNWVREN 272
Qy 239 TG 240
Db 273 SG 274

RESULT 12
S33777
hepsin (EC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: S33777; S32013
R:Farley, D.; Reymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A:Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
A:Reference number: S33777; MUID:93305733
A:Accession: S33777
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <FAR>
A:Cross-references: EMBL:X70900; NID:g57928; PIDN:CAA50256.1; PID:g57929
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:22-44/Domain: transmembrane #status predicted <TMN>
F:162-395/Domain: trypsin homology <TRY>
F:187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
F:202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 35.38; Score 466; DB 1; Length 416;
Best Local Similarity 38.3%; Pred. No. 1.6e-35;
Matches 100; Conservative 42; Mismatches 90; Indels 28; Gaps 9;

Qy 1 VVGTTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYIDDRGFRYSDPTQWTA 60
Db 162 IVGGQDSSLGRPMQVSLRYDGT-HLCGSSLGSDWLTAAHCFPE---RNRVLSRWRY 216
Qy 61 FLGLHDQSQRSAPGVQERRLKRIISH-----PFFNDFDID--DYDIALLELEKPAEYSSMV 113
Db 217 FAG---AVATSPHVAQLGVQAVIYHGSLP-FRDPTIDENSDNDIALVHLSLPLTEYI 272
Qy 114 RPICLPDASHVFPAGKATWVTGWHTQVGGTGALLQKGEIRVINQTTCE--NLLPQIT 171
Db 273 QPVCLPAGQALVDKQVTVTGWNTQFYGGQAVVLEQARVPIISNEVCNPSFYGNOIK 332
Qy 172 PRMCMVGLSGGVDSCQDSGGPL--SSVEADGRIFQAGVSWGDCQARNKPGVYTRPL 228
```

```

Db 333 PKMFCAGYPEGGIDACQDSGGHFGVCEDRISGTSRWRLGGIVSWETGCALARKPGVYTKV 392
Qy 229 PLFRDWI-----KENTGV 241
Db 393 IDFREWTFQAIKTHSEATGM 412

RESULT 13
T30337
polyprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30337
R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
submitted to the EMBL data Library, March 1998
A:Description: cDNA cloning of ovocymase, a chymotrypsin-like protease released from
A:Reference number: 220829
A:Accession: T30337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1524 <FAN>
A:Cross-references: EMBL:081290; NID:g2981640; PID:g2981641; PIDN:AAC24717.1

Query Match 35.1%; Score 463.5; DB 2; Length 1524;
Best Local Similarity 39.8%; Pred. No. 1.3e-34;
Matches 96; Conservative 42; Mismatches 92; Indels 11; Gaps 6;

Qy 1 VVGTTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYIDDRGFRYSDPTQWTA 60
Db 584 IVGGEASPNWPQVQIFLRTFTH-CEGAIISQWILTAAHC-----IRAAEPSYTWV 636
Qy 61 FLGLHDQSQRSAPGVQERRLKRIISHFFNDFDIDIALLELEKPAEYSSMVRPICLPD 120
Db 637 IAGDNRMLNEST-EQIRNIKTIRHDNANSEYDNDIALLYLEELDLNDFRVPYCLPE 695
Qy 121 ASHVFPAGKATWVTGWHTQVGGTGALLQKGEIRVINQTTCE--NLLPQITPRMCMVG 179
Db 696 PEEVLTASVCVVTGWNTAEDGQALGQLQLPLDLSICTNTSYSGELTDHMLCAG 755
Qy 180 LSG-GVDSQCDGSGPLSSVEADGRIFQAGVSWGDCQARNKPGVYTRPLFRDWIKEN 238
Db 756 PSSKEKDACQDSGGPLVCQNKEDQFSIYGLVSWGEGCGRVPYTKVRLFFFTWI-QN 814
Qy 239 T 239
Db 815 T 815

RESULT 14
S40005
trypsin (EC 3.4.21.4) precursor - African malaria mosquito
C:Species: Anopheles gambiae (African malaria mosquito)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 22-Jun-1999
C:Accession: S40005
R:Mueller, H.
submitted to the EMBL Data Library, June 1993
A:Reference number: S40003
A:Accession: S40005
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-275 <MUE>
A:Cross-references: EMBL:Z22930; NID:g410323; PIDN:CAA80515.1; PID:g410326
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase
F:49-269/Domain: trypsin homology <TRY>
F:89,134,230/Active site: His, Asp, Ser #status predicted

Query Match 35.08; Score 462; DB 2; Length 275;
Best Local Similarity 42.8%; Pred. No. 2.3e-35;
Matches 104; Conservative 30; Mismatches 91; Indels 18; Gaps 7;
```



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:10:35 ; Search time 11.03 Seconds

(without alignments)  
846.003 Million cell updates/sec

Title: US-09-657-986B-2

Perfect score: 1319

Sequence: 1 VVGGTDADEGEWPMQVSLHA.....PGVYTRLPFRDMIKENTGV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1319	100.0	855	1 ST14_HUMAN	Q9Y5Y6 homo sapien
2	1152	87.3	855	1 ST14_MOUSE	P56677 mus musculus
3	576.5	43.7	422	1 DES1_HUMAN	Q9UL52 homo sapien
4	552.5	41.9	1034	1 ENTK_PIG	P98074 sus scrofa
5	547.5	41.5	1019	1 ENTK_HUMAN	P98073 homo sapien
6	544.5	41.3	1035	1 ENTK_BOVIN	P98072 bos taurus
7	510	38.7	490	1 TMS2_MOUSE	Q9J1Q8 mus musculus
8	503	38.1	1069	1 ENTK_MOUSE	P97435 mus musculus
9	497.5	37.7	418	1 HATT_HUMAN	O60235 homo sapien
10	497	37.7	492	1 TMS2_HUMAN	O15393 homo sapien
11	483	36.6	455	1 TMS5_MOUSE	Q9ER04 mus musculus
12	483	36.6	786	1 STUB_DROME	Q03319 drosophila
13	481.5	36.5	416	1 HEP5_MOUSE	O38453 mus musculus
14	477	36.2	457	1 TMS5_HUMAN	Q9H3S3 homo sapien
15	477	36.2	638	1 KAL_MOUSE	P26262 mus musculus
16	474.5	36.0	454	1 TMS3_HUMAN	P57727 homo sapien
17	473.5	35.9	417	1 HEP5_HUMAN	P05981 homo sapien
18	472	35.8	790	1 PLMN_PIG	P06867 sus scrofa
19	471.5	35.7	290	1 MPN_HUMAN	Q9BQ33 homo sapien
20	469	35.6	638	1 KAL_RAT	P14272 rattus norv
21	467	35.4	275	1 TRY3_ANOGA	P35037 anopheles g
22	466	35.3	416	1 HEP5_RAT	Q05511 rattus norv
23	462	35.0	275	1 TRY4_ANOGA	P35038 anopheles g
24	462	35.0	638	1 KAL_HUMAN	P03952 homo sapien
25	461	35.0	343	1 PSS8_HUMAN	Q16651 homo sapien
26	460	34.9	275	1 TRY7_PIG	Q9N2D1 sus scrofa
27	457	34.6	342	1 PSS8_RAT	Q9ES87 rattus norv
28	456.5	34.6	267	1 TRY7_ANOGA	P35041 anopheles g
29	455	34.5	625	1 FALL1_HUMAN	P03951 homo sapien
30	454	34.4	274	1 TRY1_ANOGA	P35035 anopheles g
31	453.5	34.4	812	1 PLMN_MOUSE	P20918 mus musculus
32	452	34.3	343	1 PLMN_SHEEP	P81286 ovis aries
33	452	34.3	810	1 PLMN_HUMAN	P00747 homo sapien

34	448.5	34.0	810	1 PLMN_ERIEU	Q29485 erinaceus e
35	447	33.9	333	1 PLMN_CANFA	P80009 canis famil
36	446	33.8	238	1 TRY5_AEDAE	P29787 aedes aegypt
37	444	33.7	338	1 PLMN_HORSE	P80010 equus cabal
38	442	33.5	342	1 PSS8_MOUSE	Q9ESD1 mus musculus
39	441	33.4	810	1 PLMN_MACMU	P12545 macaca mula
40	441	33.4	1042	1 COR1_HUMAN	Q9Y5Q5 homo sapien
41	439.5	33.3	812	1 PLMN_BOVIN	P06868 bos taurus
42	439	33.3	270	1 TRY7_MERON	P50342 meriones un
43	437	33.1	273	1 TRY7_SHEEP	Q9XSM2 ovis aries
44	436.5	33.1	416	1 FAG_BOVIN	P00741 bos taurus
45	436	33.1	254	1 TRY3_AEDAE	P29786 aedes aegypt

ALIGNMENTS

RESULT	1
ID	ST14_HUMAN
AC	Q9Y5Y6; Q9H3S0; Q9HCA3; Q9BS01; Q9HB36; STANDARD; PRT: 855 AA.
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DE	Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matritrase) (Membrane-
DE	type serine protease 1) (MT-SPI) (Prostamin) (Serine protease TADG-15)
DE	(Tumor associated differentially-expressed gene-15 protein).
GN	ST14 OR PRSS14 OR SNC19 OR TADG15.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99303581; PubMed=10373424;
RA	Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT	"Molecular cloning of cDNA for matritrase, a matrix-degrading serine
RT	protease with trypsin-like activity.";
RL	J. Biol. Chem. 274:18231-18236(1999).
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99432178; PubMed=10500122;
RA	Takeuchi T., Shuman M.A., Craik C.S.;
RT	"Reverse biochemistry: Use of macromolecular protease inhibitors to
RT	dissect complex biological processes and identify a membrane-type
RT	serine protease in epithelial cancer and normal tissue.";
RL	Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN	[3]
RP	SEQUENCE FROM N.A.
RT	TISSUE=Prostate;
RA	Yamaguchi N., Mitsui S.;
RT	"Molecular cloning of a novel transmembrane serine protease expressed
RT	in human prostate.";
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN	[4]
RP	SEQUENCE FROM N.A.
RA	Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,
RT	O'Brien T.J.;
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN	[5]
RP	SEQUENCE OF 327-855 FROM N.A.
RA	TISSUE=Muscle;
RA	Strausberg R.;
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN	[6]
RP	SEQUENCE OF 340-664 FROM N.A.
RA	Cao J., Fan W., Zheng S.;
RT	"Genomic analysis of a novel human serine protease SNC19.";
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN	[7]
RP	CHARACTERIZATION.
RA	TISSUE=Milk;
RL	MEDLINE=99303582; PubMed=10373425;

Lin C.Y., Anders J., Johnson M., Dickson R.B.;  
 "Purification and characterization of a complex containing matrilysin  
 and a Kunitz-type serine protease inhibitor from human milk.";  
 J. Biol. Chem. 274:18237-18242(1999).  
 -!- FUNCTION: DEGRADERS EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE  
 IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE  
 ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG  
 OR LYS AS THE P1 SITE.  
 -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.  
 -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPSIN FAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AF118224; AAD42765.2; -  
 EMBL; AF113086; AAF00109.1; -  
 EMBL; AB030036; BAB20376.1; -  
 EMBL; AF057145; AAG15395.1; -  
 EMBL; BC005826; AAH05826.1; -  
 EMBL; AF283256; AAG13949.1; -  
 HSP; P00763; IDPO.  
 MEROPS; S01.302; -  
 InterPro; IPR000859; CUB.  
 InterPro; IPR001314; Chymotrypsin.  
 InterPro; IPR002172; LDL\_recept\_A.  
 InterPro; IPR001254; Trypsin.  
 Pfam; PF00431; CUB; 2; LDL\_recept\_a; 4.  
 Pfam; PF00057; LDL\_recept\_a; 4.  
 Pfam; PF00089; trypsin; 1.  
 PRINTS; PR00722; CHYMOTRYPSIN.  
 PRINTS; PR00261; LDLRECEPTOR.  
 SMART; SM00042; CUB; 2.  
 SMART; SM00192; LDLa; 3.  
 SMART; SM00020; Tryp\_SPC; 1.  
 PROSITE; PS01180; CUB; 2; 2.  
 PROSITE; PS01209; LDLRA\_1; 2.  
 PROSITE; PS00068; LDLRA\_2; 4.  
 PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 PROSITE; PS00135; TRYPSIN\_SER; 1.  
 Signal-anchor; Glycoprotein; Hydrolase; Serine protease;  
 Transmembrane; Repeat.  
 DOMAIN 1 55  
 TRANSMEM 56 76  
 CYTOPLASMIC (POTENTIAL).  
 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 (POTENTIAL).  
 EXTRACELLULAR (POTENTIAL).  
 CUB 1.  
 CUB 2.  
 LDL-RECEPTOR CLASS A 1.  
 LDL-RECEPTOR CLASS A 2.  
 LDL-RECEPTOR CLASS A 3.  
 LDL-RECEPTOR CLASS A 4.  
 SERINE PROTEASE.  
 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 PEA -> GTR (IN REF. 5).  
 R -> S (IN REF. 4).  
 A -> V (IN REF. 3).  
 855 AA; 94769 MW; 26143132C0IF99C9 CRC64;  
 SEQUENCE

Query Match 100.0%; Score 1319; DB 1; Length 855;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-117;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VVGSTDADEGEMPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTA 60  
 DB 615 VVGSTDADEGEMPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTA 674  
 QY 61 FLGLHDQSQRSAPGVQERRKRIISHFFNFDFDYDIALLELEKPAEYSMVVRPCLPD 120  
 DB 675 FLGLHDQSQRSAPGVQERRKRIISHFFNFDFDYDIALLELEKPAEYSMVVRPCLPD 734  
 QY 121 ASHVPPAGKAIWITGWHGTQYGGTGALLQKGEIRVINQTCENLLPQQLTPRMVGVFL 180  
 DB 735 ASHVPPAGKAIWITGWHGTQYGGTGALLQKGEIRVINQTCENLLPQQLTPRMVGVFL 794  
 QY 181 SGGVSCQCGDSGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRWIKENTG 240  
 DB 795 SGGVSCQCGDSGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRWIKENTG 854  
 QY 241 V 241  
 DB 855 V 855  
 RESULT 2  
 ID ST14\_MOUSE STANDARD; PRT; 855 AA.  
 AC P56677;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithin).  
 GN ST14 OR PRSS14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;  
 RX MEDLINE=99216440; PubMed=10199918;  
 RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,  
 RA Schwartz R.H.;  
 RT "Cloning and chromosomal mapping of a gene isolated from thymic  
 stromal cells encoding a new mouse type II membrane serine protease,  
 epithin, containing four LDL receptor modules and two CUB domains.";  
 RT Immunogenetics 49:420-428(1999).  
 RL [2]  
 RP REVIEWS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.  
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;  
 RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast tumor;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,  
 CC AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART,  
 CC TESTIS, AND BRAIN.  
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 CC

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)

CC EMBL; AF042822; AAO20230.3; -  
CC DR EMBL; BC005496; AAH05496.1; -  
CC DR HSSP; P20231; 1AAO.  
CC DR MEROPS; S01\_014; -  
CC DR MGD; MGI:133888; St14.  
CC DR InterPro; IPR000859; CUB.  
CC DR InterPro; IPR002172; LDL\_recept\_A.  
CC DR InterPro; IPR001254; Trypsin.  
CC DR Pfam; PF00057; ldl\_recept\_a; 4.  
CC DR Pfam; PF00089; trypsin\_1.  
CC DR PRINTS; PRO0261; LDLRECEPTOR.  
CC DR PRINTS; PRO0722; CHYMOTRYPSIN.  
CC DR SMART; SM00042; CUB; 2.  
CC DR SMART; SM00192; LDLA; 3.  
CC DR PROSITE; PS01209; CUB; 2.  
CC DR PROSITE; PS01209; LDLA; 2.  
CC DR PROSITE; PS00068; LDLA\_2; 4.  
CC DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
CC KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;  
KW Transmembrane; Repeat.  
FT DOMAIN 1 55  
FT TRANSMEM 56 76  
FT  
FT DOMAIN 77 855  
FT ACT\_SITE 214 331  
FT DOMAIN 340 444  
FT DOMAIN 451 488  
FT DOMAIN 489 522  
FT DOMAIN 523 561  
FT DOMAIN 565 604  
FT DOMAIN 615 854  
FT ACT\_SITE 656 656  
FT ACT\_SITE 711 711  
FT ACT\_SITE 805 805  
FT CARBOHYD 107 107  
FT CARBOHYD 302 302  
FT CARBOHYD 365 365  
FT CARBOHYD 421 421  
FT CARBOHYD 489 489  
FT CARBOHYD 772 772  
SQ SEQUENCE 855 AA; 94654 MW; 4F10E84DA2146DD5 CRC64;

Query Match 87.3%; Score 1152; DB 1; Length 855;

Best Local Similarity 86.3%; Pred. No. 3.8e-101;  
Matches 208; Conservative 19; Mismatches 14; Indels 0; Gaps 0;

QY 1 VVGTDADGEGPQVSLHALGQGHICGASLISPNWLYSAARHCYIDDRGFRYSPTQWTA 60  
Db 615 VVGGINADGEGPQVSLHALGQGHICGASLISPNWLYSAARHCYIDDRGFRYSPTQWTA 674  
QY 61 FLGLHDQSORAPGVQERRLKRITRSHPFNFDTFDYDIALLEKPEYSSMVRPICLPD 120  
Db 675 FLGLHDQSORAPGVQERRLKRITRSHPFNFDTFDYDIALLEKPEYSSMVRPICLPD 734  
QY 121 ASHVPFAGKAIWVTGWTGVTGGLTQGETRVINOTTENLLPQQTIPRMVGVFL 180  
Db 735 AHTVPFAGKAIWVTGWTGVTGGLTQGETRVINOTTENLLPQQTIPRMVGVFL 794  
QY 181 SGGVDSQSGDGGPLSSVDEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENTG 240  
Db 795 SGGVDSQSGDGGPLSSVDEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENTG 854  
QY 241 V 241  
Db 855 V 855

RESULT 3  
DESI\_HUMAN STANDARD; PRT; 422 AA.  
AC Q9UL52;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Serine protease DESCI precursor (EC 3.4.21.-).  
GN DESCI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21094880; PubMed=11161383;  
RA Lang J.C., Schuller D.E.;  
RT "Differential expression of a novel serine protease homologue in  
RT squamous cell carcinoma of the head and neck";  
EL Br. J. Cancer 84:237-243(2001).  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY  
CC CLEAVAGE AND SECRETED (By similarity).  
CC -!- TISSUE SPECIFICITY: Expression can only be detected in tissues  
CC derived from the head and neck, and in skin, prostate and testis.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
CC  
CC EMBL; AF064819; AAF04328.1; -  
CC HSSP; P00763; LDPO.  
CC MEROPS; S01\_021; -  
CC InterPro; IPR001314; Chymotrypsin.  
CC InterPro; IPR000082; SEA.  
CC InterPro; IPR001254; Trypsin.  
CC Pfam; PF01390; SEA; FALSE\_NEG.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PRO0722; CHYMOTRYPSIN.  
CC SMART; SM00200; SEA; 1.  
CC SMART; SM00020; Tryp\_Spc; 1.  
CC PROSITE; PS00024; SEA; 1.  
CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;  
KW Glycoprotein.  
CHAIN 1 190 SERINE PROTEASE DESCI, NON-CATALYTIC  
CHAIN (POTENTIAL).  
CHAIN 191 422 SERINE PROTEASE DESCI, CATALYTIC CHAIN  
(POTENTIAL).  
DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).  
TRANSMEM 19 39 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
DOMAIN 40 422 EXTRACELLULAR (POTENTIAL).  
DOMAIN 43 164 SEA.  
DOMAIN 191 421 SERINE PROTEASE.  
ACT\_SITE 231 231 CHARGE RELAY SYSTEM (BY SIMILARITY).  
ACT\_SITE 276 276 CHARGE RELAY SYSTEM (BY SIMILARITY).  
DISULFID 372 372 CHARGE RELAY SYSTEM (BY SIMILARITY).  
DISULFID 176 296 INTERCHAIN (BY SIMILARITY).  
DISULFID 216 232 BY SIMILARITY.  
DISULFID 341 357 BY SIMILARITY.  
DISULFID 368 397 BY SIMILARITY.  
CARBOHYD 74 74 N-LINKED (GLCNAC...) (POTENTIAL).  
CARBOHYD 165 165 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 222 222 N-LINKED (GLNAC, . . .) (POTENTIAL).  
SQ SEQUENCE 422 AA; 47564 MW; 595600F944FF9494 CRC64;

Query Match 43.7%; Score 576.5; DB 1; Length 422;

Best Local Similarity 46.9%; Pred. No. 4.4e-47;  
Matches 115; Conservative 36; Mismatches 77; Indels 17; Gaps 5;

QY 1 VVGTDADGEWQVSHALQGHGICGASLISPNWLSAHCYIDRGRYSDPTQWTA 60  
DB 191 IVGGTEVEGEPWQASLOWQD -SHRCGATLINATWLSAHCFT-----TYKNPARWTA 244  
QY 61 FLGLHDQSORSAPGVOERLRKRRIISHPFFNDFTDYDIALLEKPAEYSSMVRPCLPD 120  
DB 245 SFGV-----TIKPSKMKRGLRIIVHEKYKHPSHDYDISLAELSSPPVYTNAVHVRCLPD 299  
QY 121 ASHVFPAGKAIWITGWHTQVGGTGALLIQGEIRVINQITCENLLPQ----QITPRMCM 176  
DB 300 ASYEPQGDVMEVTFGALKNDGYSNHLRQAVTLIDATCNE--FOAYNDAITPRMCLC 357  
QY 177 VGFLEGGVDCOGDGGPLSSVEADGRIFQAGVYSWGCGAQRNKPQVYTRLPFRDWIK 236  
DB 358 ASLSEKTDACGDSGGPLVSSDARDIWLAGIYSWGDECAKPNKPGVYTRYTALRDWIT 417  
QY 237 ENTGV 241  
DB 418 SKTGI 422

## RESULT 4

ENTK\_PIG STANDARD; PRT: 1034 AA.

AC P98074;  
DT 01-FEB-1996 (Rel. 33; Created)  
DT 01-FEB-1996 (Rel. 33; Last sequence update)  
DT 16-OCT-2001 (Rel. 40; Last annotation update)  
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).  
GN PRSS7 OR ENTK.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;

[1]  
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=DuoDenal mucosa;

RX MEDLINE=94327348; PubMed=8051081;

RA Matsushima M., Ichinose M., Yanagi N., Kakei N., Tsukada S.,

RA Miki K., Kurokawa K., Tashiro K., Shikawa K., Shinomiya K.,

RA Uneyama H., Inoue H., Takahashi T., Takahashi K.;

RT "Structural characterization of porcine enteropeptidase.";

RL J. Biol. Chem. 269:19976-19982(1994).

CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC

CC PROTEOLYTIC PROENZYME (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE

CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN

CC TURN ACTIVATES OTHER PROENZYME INCLUDING CHYMOTRYPSINOGEN,

CC PROCARBOXYPEPTIDASES, AND PROELASTASES.

CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in

CC trypsinogen.

CC -1- SUBUNIT: HETEROTRIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN

CC (HEAVY) CHAIN, AND A MINI CHAIN.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).

CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS

CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.

CC -1- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.

CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; D30799; BAA06459.1; -.  
DR HSSP; P20231; 1AAO.  
DR MEROPS; S01.156; -.  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR002172; LDL\_recept\_A.  
DR InterPro; IPR000598; MAM.  
DR InterPro; IPR000082; SEA.  
DR InterPro; IPR001190; SRCR.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00431; CUB; 2.  
DR Pfam; PF00057; ldl\_recept\_a; 2.  
DR Pfam; PF00629; MAM; 1.  
DR Pfam; PF01390; SEA; 1.  
DR Pfam; PF00530; SRCR; 1.  
DR Pfam; PF00089; Trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00042; CUB; 2.  
DR SMART; SM00192; LDLa; 2.  
DR SMART; SM00137; MAM; 1.  
DR SMART; SM00200; SEA; 1.  
DR SMART; SM00202; SR; 1.  
DR SMART; SM00020; Tryp\_spc; 1.  
DR PROSITE; PS01180; CUB; 2.  
DR PROSITE; PS01209; LDLRA\_1; 2.  
DR PROSITE; PS00068; LDLRA\_2; 2.  
DR PROSITE; PS00740; MAM\_1; 1.  
DR PROSITE; PS00060; MAM\_2; 1.  
DR PROSITE; PS00024; SEA; 1.  
DR PROSITE; PS00420; SRCR\_1; FALSE\_NEG.  
DR PROSITE; PS0287; SRCR\_2; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;  
KW Serine protease; Zymogen; Transmembrane; Repeat.  
FT CHAIN 52 117 NON-CATALYTIC M CHAIN (MINI CHAIN).  
FT CHAIN 118 799 NON-CATALYTIC H CHAIN (HEAVY CHAIN).  
FT CHAIN 800 1034 CATALYTIC L CHAIN (LIGHT CHAIN).  
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
FT DOMAIN 48 1034 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 52 169 SEA.  
FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.  
FT DOMAIN 240 349 CUB 1.  
FT DOMAIN 357 519 MAM.  
FT DOMAIN 539 649 CUB 2.  
FT DOMAIN 656 694 LDL-RECEPTOR CLASS A 2.  
FT DOMAIN 693 786 SRCR.  
FT DOMAIN 800 1034 SERINE PROTEASE.  
FT ACT\_SITE 840 840 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 891 891 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 986 986 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT LIPID 2 2 MYRISTATE (POTENTIAL).  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 206 225 BY SIMILARITY.  
FT DISULFID 219 236 BY SIMILARITY.  
FT DISULFID 638 670 BY SIMILARITY.  
FT DISULFID 665 683 BY SIMILARITY.  
FT DISULFID 677 692 BY SIMILARITY.  
FT DISULFID 787 911 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 825 841 BY SIMILARITY.  
FT DISULFID 925 952 BY SIMILARITY.  
FT DISULFID 956 971 BY SIMILARITY.  
FT DISULFID 982 1010 BY SIMILARITY.



```

FT CARBOHYD 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 549 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 697 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 701 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 740 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 761 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 804 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 863 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1034 AA; 114775 MB; 0388C64CF64CC368 CRC64;

Query Match 41.9%; Score 552.5; DB 1; Length 1034;
Best Local Similarity 44.7%; Pred. No. 2.3e-44;
Matches 106; Conservative 45; Mismatches 79; Indels 7; Gaps 4;

QY 1 VVGTDADSGEPWQVSLHALGGHICGASLISPNWLSAAHCVIDDRGFYSDPTQWTA 60
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 800 IGVGNDREGAPWVVALYNGQ-LICGASLSRDLVLSAAHCVVG-----RNLEPSRWKA 854
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 FLGLHQDSQAPGVOERKLRIISHPFNDFDYDIALLEKPAEYSWVRPICLPD 120
   ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 855 ILGLHMTSLSPQIVRLIDEIVNPHVNRKDSDIAMHLEFKVNYTDYIQICLPE 914
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 ASHVPAGTAIVTGWHTQYGGTGAIILOKGEIRIVNQTICENLLPQ-QITPRMCMGVF 179
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 915 ENGVFPFGRICIAGKVIYQGSFADILQEADEPLLSNERKCCQQQMPYNTENMCAGY 974
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 180 LSGGVDSQDGGSLPSSVEADGRIFQAGVSWGDCGAKRKPQVYTRPLFRDWIK 236
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 975 EEGIIDSQDGGSLPCLMLE-NRWLLAGVTSFGYQCCLPRLPQVYARVPKFTWQI 1030
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
ENTK_HUMAN STANDARD; PRT; 1019 AA.
ID ENTK_HUMAN AC P98073;
DC 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN PRSS7 OR ENTK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Duodenum;
RX MEDLINE=95234679; PubMed=7718557;
RA Kitamoto Y., Veille R.A., Donis-Keller H., Sadler J.E.;
RT "cDNA sequence and chromosomal localization of human enterokinase,
RL the proteolytic activator of trypsinogen.";
RN Biochemistry 34:4562-4568(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Holzinger A., Buck C., Maier E.M., Mayerhofer P.U., Roscher A.A.,
RA Sadler J.E., Hadorn H.B.;
RT "Genomic organization of the human enteropeptidase.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

```

```

[3]
SEQUENCE FROM N.A.
MEDLINE=20289799; PubMed=10830953;
RA Park H.-S., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Hattori M., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordliek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Degand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yasuo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
[4]
SEQUENCE OF 749-1019 FROM N.A.
TISSUE=Duodenum;
MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
CC trypsinogen.
CC -!- SUBUNIT: HETEROIDIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -!- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
CC MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -!- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 NAM DOMAIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U09660; AAC50138.1;
EMBL; Y19124; CAB65555.1;
EMBL; Y19125; CAB65555.1; JOINED.
EMBL; Y19126; CAB65555.1; JOINED.
EMBL; Y19127; CAB65555.1; JOINED.
EMBL; Y19128; CAB65555.1; JOINED.
EMBL; Y19129; CAB65555.1; JOINED.
EMBL; Y19130; CAB65555.1; JOINED.
EMBL; Y19131; CAB65555.1; JOINED.
EMBL; Y19132; CAB65555.1; JOINED.
EMBL; Y19133; CAB65555.1; JOINED.
EMBL; Y19134; CAB65555.1; JOINED.
EMBL; Y19135; CAB65555.1; JOINED.
EMBL; Y19136; CAB65555.1; JOINED.
EMBL; Y19137; CAB65555.1; JOINED.
EMBL; Y19138; CAB65555.1; JOINED.
EMBL; Y19139; CAB65555.1; JOINED.

```

DR EMBL; Y19140; CAB65555.1; JOINED.  
 DR EMBL; Y19141; CAB65555.1; JOINED.  
 DR EMBL; Y19142; CAB65555.1; JOINED.  
 DR EMBL; Y19143; CAB65555.1; JOINED.  
 DR EMBL; AL163218; CAB90392.1; .  
 DR EMBL; AL163217; CAB90389.1; .  
 DR HSP; P00763; LDPO.  
 DR MEROPS; S01.156; .  
 DR MIM; 226200; .  
 DR InterPro; IPRO00859; CUB.  
 DR InterPro; IPRO01314; Chymotrypsin.  
 DR InterPro; IPRO01172; LDL\_recept\_A.  
 DR InterPro; IPRO00998; MAM.  
 DR InterPro; IPRO00082; SEA.  
 DR InterPro; IPRO01190; SRCR.  
 DR InterPro; IPRO01254; Trypsin.  
 DR Pfam; PF00431; CUB; 2.  
 DR Pfam; PF00057; ldl\_recept\_a; 2.  
 DR Pfam; PF00629; MAM; 1.  
 DR Pfam; PF01390; SEA; 1.  
 DR Pfam; PF00530; SRCR; 1.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00042; CUB; 2.  
 DR SMART; SM00132; LDLA; 2.  
 DR SMART; SM00137; MAM; 1.  
 DR SMART; SM00200; SEA; 1.  
 DR SMART; SM00202; SR; 1.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS01209; LDLA\_1; 2.  
 DR PROSITE; PS00068; LDLA\_2; 2.  
 DR PROSITE; PS00740; MAM\_1; 1.  
 DR PROSITE; PS00060; MAM\_2; 1.  
 DR PROSITE; PS00024; SEA; 1.  
 DR PROSITE; PS00420; SRCR\_1; FALSE\_NEG.  
 DR PROSITE; PS00287; SRCR\_2; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Signal-anchor: Glycoprotein; Myristate; Hydrolase;  
 Serine protease; Zymogen; Transmembrane; Repeat.  
 FT CHAIN 1 784  
 FT CHAIN 785 1019  
 FT DOMAIN 1 18  
 FT TRANSMEM 19 47  
 FT DOMAIN 48 1019  
 FT DOMAIN 52 169  
 FT DOMAIN 182 223  
 FT DOMAIN 225 334  
 FT DOMAIN 342 504  
 FT DOMAIN 524 634  
 FT DOMAIN 641 679  
 FT DOMAIN 678 771  
 FT DOMAIN 785 1019  
 FT ACT\_SITE 825 825  
 FT ACT\_SITE 876 876  
 FT ACT\_SITE 971 971  
 FT LIPID 2 2  
 FT DISULFID 184 197  
 FT DISULFID 191 210  
 FT DISULFID 204 221  
 FT DISULFID 643 655  
 FT DISULFID 650 668  
 FT DISULFID 662 677  
 FT DISULFID 772 896  
 FT DISULFID 810 826  
 FT DISULFID 910 977  
 FT DISULFID 941 956  
 FT DISULFID 967 995  
 FT CARBOHYD 116 116  
 FT CARBOHYD 147 147

FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 134 134 Q -> E (IN REF. 3).  
 FT CONFLICT 732 732 S -> P (IN REF. 3).  
 FT CONFLICT 754 771 SQCLQDSLRLQCNEHS -> RRNAKNEIDALSPIILIA (IN REF. 3).  
 SQ SEQUENCE 1019 AA; 112923 MW; B6AAA245F6DA563 CRC64;  
 Query Match 41.5%; Score 547.5; DB 1; Length 1019;  
 Best Local Similarity 44.7%; Pred. No. 6.8e-44;  
 Matches 106; Conservative 44; Mismatches 80; Indels 7; Gaps 4;  
 QY 1 VVGTDADGEWPQVSLHALGQHGICGASLISPNWLVSAAHCVIDDRGFRYSDPTQWTA 60  
 DB 785 IVGGSNAKEAGWPVVGILY-YGRLLCGASLVSSDWLVSAAHCVYG----RNLEPSKWT 839  
 QY 61 FLGLHDQSRSPAGVQVRRRLKRIISHPFNFDFDYDIALLELEKPAEYSSMYRPICLPD 120  
 DB 840 ILGLHKNLSPTQTPRLIDEIVINPHNRRKNDNDIAMHLEFKVNYTDYIQPICLPE 899  
 QY 121 ASHVPFPAKAIWVTGWGHTQYGGTGALILQGEIRVNOTTCENLLPQ-QITPRMVCVF 179  
 DB 900 ENQVPPPGNCSIAAGTVYQGITAILQEAADVPLUSNERCQQQOMPEYNITENMICAGY 959  
 QY 180 LSGVDSQCGSGGLSSVEADGRIFQAGVYVSMGDSGAQRNKFVYVTRLPFLFRDWIK 236  
 DB 960 BEGGIDSQCGSGGLMCQE-NNRWFLAGVTSGYKCALPNRPVGVARVSRFTEWIQ 1015  
 RESULT 6  
 ENTK\_BOVIN  
 ID ENTK\_BOVIN STANDARD; PRT; 1035 AA.  
 AC P98072;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).  
 GN PRS7 OR ENTK.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_taxid=9913;  
 RN [1]  
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX TISSUE=Duodenum;  
 RX MEDLINE=94329561; PubMed=8052624;  
 RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;  
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic  
 protease composed of a distinctive assortment of domains";  
 Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).  
 RL [2]  
 RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=94043122; PubMed=8226855;  
 RA Lavallie E.R., Rehemtulla A., Racie L.A., Diblasio E.A.,  
 RA Ferenc C., Grant K.L., Light A., McCoy J.M.;  
 RT "Cloning and functional expression of a cDNA encoding the catalytic  
 subunit of bovine enterokinase";

J. Biol. Chem. 266:23311-23317(1993).  
[3]  
SEQUENCE OF 801-827.  
Tissue=Intestine;  
MEDLINE=92189715; PubMed=1799406;  
Light A., Janska H.;  
"The amino-terminal sequence of the catalytic subunit of bovine  
enterokinase.";  
J. Protein Chem. 10:475-480(1991).  
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC  
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE  
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN  
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,  
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.  
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in  
CC trypsinogen.  
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A  
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.  
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS  
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.  
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; U09859; AAB40026.1; -;  
DR EMBL; L19663; AAL16035.1; -;  
DR PIR; A61436; A61436.  
DR HSSP; P20231; LAO.  
DR MEROPS; S01.156;  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR002172; LDL\_recept\_A.  
DR InterPro; IPR000998; MAM.  
DR InterPro; IPR000082; SEA.  
DR InterPro; IPR001190; SRCR.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00431; CUB; 2.  
DR Pfam; PF00057; ldl\_recept\_a; 2.  
DR Pfam; PF00629; MAM; 1.  
DR Pfam; PF01390; SEA; 1.  
DR Pfam; PF00530; SRCR; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00042; CUB; 2.  
DR SMART; SM00192; LDLA; 2.  
DR SMART; SM00137; MAM; 1.  
DR SMART; SM00200; SEA; 1.  
DR SMART; SM00202; SR; 1.  
DR SMART; SM00020; TRYP\_SPC; 1.  
DR PROSITE; PS01180; CUB; 2.  
DR PROSITE; PS01209; LDLA\_1; 2.  
DR PROSITE; PS00068; LDLA\_2; 2.  
DR PROSITE; PS00740; MAM\_1; 1.  
DR PROSITE; PS00060; MAM\_2; 1.  
DR PROSITE; PS50024; SEA; 1.  
DR PROSITE; PS00420; SRCR\_1; FALSE\_NEG.  
DR PROSITE; PS50287; SRCR\_2; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;  
KW Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.  
FT CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).  
FT CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).  
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 48 1035 SEA.  
FT DOMAIN 54 169 LDL-RECEPTOR CLASS A 1.  
FT DOMAIN 197 238 CUB 1.  
FT DOMAIN 240 350 MAM.  
FT DOMAIN 358 520 CUB 2.  
FT DOMAIN 540 650 SRCR.  
FT DOMAIN 657 695 SERINE PROTEASE.  
FT DOMAIN 694 787 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DOMAIN 801 1035 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 987 987 MYRISTATE (POTENTIAL).  
FT LIPID 2 2 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 206 225 BY SIMILARITY.  
FT DISULFID 219 236 BY SIMILARITY.  
FT DISULFID 659 671 BY SIMILARITY.  
FT DISULFID 666 684 BY SIMILARITY.  
FT DISULFID 678 693 BY SIMILARITY.  
FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 826 842 BY SIMILARITY.  
FT DISULFID 926 993 BY SIMILARITY.  
FT DISULFID 957 972 BY SIMILARITY.  
FT DISULFID 983 1011 BY SIMILARITY.  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 147 147 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 194 194 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 263 263 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 264 264 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 486 486 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 646 646 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 722 722 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 762 762 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 864 864 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 903 903 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARSPLIC 166 192 MISSING (IN SHORT ISOFORM).  
FT CONFLICT 808 808 R -> Y (IN REF. 3).  
SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

Query Match 41.3%; Score 544.5; DB 1; Length 1035;  
Best Local Similarity 43.5%; Pred. No. 1.3e-43;  
Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 1 VVGGTDADEGEWPFQVSLHALGQGTGASLISPNWLYSAAHCYIDDRGFYSPTQWTA 60  
Db 801 IVGGSDREGANFWVVALYFDQ-QVCGASLSVRDLVLSAAHCYVG----RNMEPSKWA 855  
QY 61 FLGLHDQSQRSPAGVQERKLRIISHPFNDFTFDYDIALLELEKPAEYSVMRPICLP 120  
Db 856 VLGLHMASNLTSPIQETRLDQIVINPHYNKRKKNNDIAMHLEMKVNYTDYIQICLPE 915  
QY 121 ASHVFPAGKAIWVTGWHGTQYGGTCLILORGEIRVINTTCENLLPQ-QITPRMCKVGF 179  
Db 916 ENQVFPFGRICTIAGWGALIIYQGSTADVLQEAADVPLLSKCKQOOMPENITENNVKAGY 975

```
QY 180 LSGVDSQSGSGGLPSVSEADGRIFQAGVYVSGDGAQNRKQGVYTRLPFRDWIK 236
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 976 EAGGVDSQSGSGGLPMQCE-NNRWLLAGVTSTFGYQCALPFRGVARVPRFTWFIQ 1031
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
TMS2_MOUSE
ID TMS2_MOUSE STANDARD; PRT; 490 AA.
AC Q9J1Q8; O5JKC4; O9CY82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
GN transmembrane protein X).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALB/C;
RA MEDLINE=21104370; PubMed=11169526;
RX Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and
human tissues.";
RL J. Pathol. 193:134-140(2001).
RN [2]
SEQUENCE FROM N.A.
RP Han J., Kim S.;
RT "Putative transmembrane protease X";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT "A novel mosaic serine protease, epitheliasin.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CLAVAGE AND SECRETED (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF199362; AAF97867.1; -.
CC EMBL; AF243500; AAF64186.1; -.
CC EMBL; AF113596; AAF21308.1; -.
CC HSSP; P00761; IAKS.
CC MGD; MGI:1354381; Tmprss2.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR002172; LDL_recept_LA.
CC InterPro; IPR001190; SRCR.
CC Pfam; PF00057; ldl_recept_a.1.
CC Pfam; PF00089; trypsin.1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00192; LDLa.1.
CC SMART; SM00202; SR; 1.
CC SMART; SM00020; Tryp-Src; 1.
CC PROSITE; PS01209; LDLRA.1; 1.
CC PROSITE; PS00068; LDLRA.2; 1.
CC PROSITE; PS00420; SRCR.1; FALSE_NEG.
CC PROSITE; PS00287; SRCR.2; 1.
```

```
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
KW HYDROLASE; SERINE PROTEASE; Transmembrane; Signal-anchor; Zymogen.
FT CHAIN 1 253
FT TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT CATALYTIC CHAIN.
FT CHAIN 254 490
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CYTOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 105 490
FT DOMAIN 111 149
FT DOMAIN 150 242
FT DOMAIN 254 490
FT ACT_SITE 294 294
FT ACT_SITE 343 343
FT ACT_SITE 439 439
FT SITE 253 254
FT DISULFID 76 125
FT DISULFID 119 138
FT DISULFID 132 147
FT DISULFID 171 230
FT DISULFID 184 240
FT DISULFID 243 363
FT DISULFID 279 295
FT DISULFID 408 424
FT DISULFID 435 463
FT CARBOHYD 111 111
FT CARBOHYD 212 212
FT CARBOHYD 474 474
FT CONFLICT 122 122
FT CONFLICT 178 178
FT CONFLICT 320 320
FT CONFLICT 474 474
SQ SEQUENCE 490 AA; 53479 MW; 07D2B03EA4D8A1A9 CRC64;

Query Match 38.78; Score 510; DB 1; Length 490;
Best Local Similarity 40.94; Pred. No. 1e-40;
Matches 97; Conservative 48; Mismatches 82; Indels 10; Gaps 5;

QY 1 VVGSDADGEGWPQVSLHALGQGHICGASLISPNVLVSAAHGVYDIDRGFRYSDPTQWTA 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 254 IVGGLNASPGDMPQVSLHVQGV-HVCGSGIITPEWIVTAHCVVEP---LSGPRYWTA 308

QY 61 FLGLHDQSQRSAPGVQERRLKRIISHPFNDFTDYDIALLELEKPAEYSSMYRPLCLPD 120
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 309 FAGILRQSLMEFYG--SRHQVEKVISHPNYSKTKNNDIALMKLQTLPLAFNDLVKPYCLPN 366

QY 121 ASHVFPAGKAIWVTGWGHTGYGGTGALILQKGEIRVINTTCEN--LLPQOITPRMVCVG 178
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 FGMMLDLDDQECWISGWGATYKGTSDVLAAMVPLLEPSKCNKSKYIYNNLIITPAMICAG 426

QY 179 FLISGVDSQSGSGGLPSVSEADGRIFQAGVYVSGDGAQNRKQGVYTRLPFRDWI 235
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 427 FLQSGVDSQSGSGGLVTLK-NGIWWLLIGTWSGSCAKALRPGVGNVTVFTDWI 482

RESULT 8
ENTK_MOUSE
ID ENTK_MOUSE STANDARD; PRT; 1069 AA.
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enteropeptidase (EC 3.4.21.9) (Enterokinase).
GN PRS7 OR ENTK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
```

SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Duoenum;  
 RX MEDLINE=98147142; PubMed=9486188;  
 RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;  
 RT "Structure of murine enterokinase (enteropeptidase) and expression in  
 RL small intestine during development."; Am. J. Physiol. 274:G342-G349(1998).  
 CC -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC  
 CC PROTEOLYTIC PREENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE  
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN  
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,  
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in  
 CC trypsinogen.  
 CC -!- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A  
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -!- PM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS  
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U73378; AAB37317.1; -.  
 CC HSP: P20231; IAAO.  
 CC MEROPS: S01.156; -.  
 CC MGD: MGI:1197523; Prss7.  
 CC InterPro: IPR000859; CUB.  
 CC InterPro: IPR001334; Chymotrypsin.  
 CC InterPro: IPR002172; LDL\_recept\_A.  
 CC InterPro: IPR000998; MAM.  
 CC InterPro: IPR000082; SEA.  
 CC InterPro: IPR001190; SRCR.  
 CC InterPro: IPR001254; Trypsin.  
 CC Pfam: PF00431; CUB; 2.  
 CC Pfam: PF00057; ldl\_recept\_a; 2.  
 CC Pfam: PF00629; MAM; 1.  
 CC Pfam: PF01390; SEA; 1.  
 CC Pfam: PF00530; SRCR; 1.  
 CC Pfam: PF00089; trypsin; 1.  
 CC PRINTS: PR00722; CHYMOTRYPSIN.  
 CC SMART: SM00042; CUB; 2.  
 CC SMART: SM00192; LDLA; 2.  
 CC SMART: SM00137; MAM; 1.  
 CC SMART: SM00200; SEA; 1.  
 CC SMART: SM00202; SR; 1.  
 CC SMART: SM00020; Tryp\_SPC; 1.  
 CC PROSITE: PS01180; CUB; 2.  
 CC PROSITE: PS01209; LDLRA\_1; 2.  
 CC PROSITE: PS00088; LDLRA\_2; 2.  
 CC PROSITE: PS00740; MAM\_1; 1.  
 CC PROSITE: PS00740; MAM\_2; 1.  
 CC PROSITE: PS50060; MAM; 1.  
 CC PROSITE: PS50024; SEA; 1.  
 CC PROSITE: PS00420; SRCR\_1; FALSE\_NEG.  
 CC PROSITE: PS0287; SRCR\_2; 1.  
 CC PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 CC PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE: PS00135; TRYPSIN\_SER; 1.  
 CC Signal-anchor; Glycoprotein; Myristate; Hydrolase;  
 KW Serine protease; zymogen; Transmembrane; Repeat.

PT	CHAIN	1	829	NON-CATALYTIC CHAIN (HEAVY CHAIN).
PT	CHAIN	830	1069	CATALYTIC CHAIN (LIGHT CHAIN).
PT	DOMAIN	1	18	CITOPASMIC (POTENTIAL).
PT	TRANSMEM	19	47	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
PT	DOMAIN	48	1069	EXTRACELLULAR (POTENTIAL).
PT	DOMAIN	52	169	SEA.
PT	DOMAIN	227	268	LDL-RECEPTOR CLASS A 1.
PT	DOMAIN	270	379	CUB 1.
PT	DOMAIN	387	549	MAM.
PT	DOMAIN	569	679	CUB 2.
PT	DOMAIN	686	724	LDL-RECEPTOR CLASS A 2.
PT	DOMAIN	723	816	SRCR.
PT	DOMAIN	830	1069	SERINE PROTEASE.
PT	ACT_SITE	874	874	CHARGE RELAY SYSTEM (BY SIMILARITY).
PT	ACT_SITE	925	925	CHARGE RELAY SYSTEM (BY SIMILARITY).
PT	ACT_SITE	1021	1021	CHARGE RELAY SYSTEM (BY SIMILARITY).
PT	LIPID	2	2	MYRISTATE (POTENTIAL).
PT	DISULFID	229	242	BY SIMILARITY.
PT	DISULFID	236	255	BY SIMILARITY.
PT	DISULFID	249	266	BY SIMILARITY.
PT	DISULFID	688	700	BY SIMILARITY.
PT	DISULFID	695	713	BY SIMILARITY.
PT	DISULFID	707	722	BY SIMILARITY.
PT	DISULFID	817	945	INTERCHAIN (BY SIMILARITY).
PT	DISULFID	859	875	BY SIMILARITY.
PT	DISULFID	959	1027	BY SIMILARITY.
PT	DISULFID	991	1006	BY SIMILARITY.
PT	DISULFID	1017	1045	BY SIMILARITY.
PT	CARBOHYD	147	147	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	197	197	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	212	212	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	373	373	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	380	380	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	433	433	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	515	515	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	579	579	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	675	675	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	727	727	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	751	751	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	770	770	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	791	791	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	897	897	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	936	936	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	999	999	N-LINKED (GLCNAC. .) (POTENTIAL).
SEQ	SEQUENCE	1069 AA;	118735 MW;	E52549E463743C3D CRC64;

Query Match 38.1%; Score 503; DB 1; Length 1069;  
 Best Local Similarity 42.1%; Pred. No. 1.2e-39;  
 Matches 101; Conservative 46; Mismatches 83; Indels 10; Gaps 5;

QY	1	VVGSTDADEGEWPOVSLHALGQGH--ICGASLISPNLVSAAHCYIDDRGRYSDPTQ 57
Db	830	IVGSDAQAGANFWVALYHRDSTRDLCLGASLVSDWLVSAAHCVYR----RNLDPTR 885
QY	58	WTAFLGLHDQSRSAFGVQERRLKRRIISHPPFFNDFTDYDIALLELEKPAEYSSMVRPIC 117
Db	886	WTAFLGLHMQSNLTSPOVRRVVDQIVINPHVDRRRKVNDIAMHLEFKVNTDYIQPIC 945
QY	118	LPDASHVFPAGKAIWWTGCHTQY-GTGCALILQKEIRVINQTTCCENLLPQ-QITPRMK 175
Db	946	LPEENQIFIPGRTCSTAGWGYDKINAGSTVDVLKADVPLISNEKCCQQLPEYNITESMI 1005
QY	176	CVGFLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQACQAKNKPQVYTRLPFLFRDWI 235
Db	1006	CAGVEEGIDSCQDGGPLMCQE-NRWFLVGVTSFGVQCCLPNHPGVYVRVSQFIEMI 1064

RESULT 9  
 HATT\_HUMAN  
 ID HATT\_HUMAN  
 AC O60235; STANDARD; PRT; 418 AA.

DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Airway trypsin-like protease precursor (EC 3.4.21.-).  
HA1.  
GN Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed=9565616;  
RA MEDLINE=98234382; Masuda K.-I., Ogawa H., Takagi K., Umemoto N., Yasuoka S.;  
RA Yasuoka K., Masuda K.-I., Ogawa H., Takagi K., Umemoto N., Yasuoka S.;  
RT "Cloning and characterization of the cDNA for human airway trypsin-  
RT like protease";  
RL J. Biol. Chem. 273:11895-11901(1998).  
RN [2]  
RP SEQUENCE OF 187-206, AND CHARACTERIZATION.  
RX MEDLINE=97224034; PubMed=9070615;  
RA Yasuoka S., Ohnishi T., Kawano S., Tsuchihashi S., Ogawara M.,  
RA Masuda K.-I., Yamaoka K., Takahashi M., Sano T.;  
RT "Purification, characterization, and localization of a novel  
RT trypsin-like protease found in the human airway";  
RL Am. J. Respir. Cell Mol. Biol. 15:300-308(1997).  
CC -!- FUNCTION: May play some biological role in the host defense system  
CC on the mucous membrane independently of or in cooperation with  
CC other substances in airway mucous or bronchial secretions.  
CC -!- CATALYTIC ACTIVITY: Preferentially cleaves the C-terminal side of  
CC arginine residues at the P1 position of certain peptides, cleaving  
CC Boc-Phe-Ser-Arg-4-methylcoumaryl-7-amide most efficiently and  
CC having an optimum pH of 8.6 with this substrate.  
CC -!- ENZYME REGULATION: Strongly inhibited by diisopropyl  
CC fluorophosphate, leupeptin, antipain, aprotinin, and soybean  
CC trypsin inhibitor, but hardly inhibited by secretory leukocyte  
CC protease inhibitor at 10 microm.  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY  
CC CLEAVAGE AND SECRETED.  
CC -!- TISSUE SPECIFICITY: Located in the cells of the submucosal serous  
CC glands of the bronchi and trachea.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB02134; BAA28691.1; -.  
CC HSP; P00750; IRTF.  
CC MIM; 605369; -.  
CC MEROPS; S01.301; -.  
CC InterPro; IPR001314; Chymotrypsin.  
CC InterPro; IPR000082; SEA.  
CC InterPro; IPR001254; Trypsin.  
CC Pfam; PF01390; SEA; 1.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC SMART; SM00200; SEA; 1.  
CC SMART; SM00020; TRYP\_SPC; 1.  
CC PROSITE; PS00024; SEA; 1.  
CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
CC Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;  
KW Glycoprotein.  
FT CHAIN 1 186 AIRWAY TRYPSIN-LIKE PROTEASE, NON-  
FT CATALYTIC CHAIN.  
FT CHAIN 187 418 AIRWAY TRYPSIN-LIKE PROTEASE, CATALYTIC

FT DOMAIN 1 20 CHAIN.  
FT TRANSMEM 21 41 CYTOPLASMIC (POTENTIAL).  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL)  
FT EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 42 418 SEA.  
FT DOMAIN 44 164 SERINE PROTEASE.  
FT ACT\_SITE 187 417 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 227 272 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 272 272 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 368 368 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 173 292 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 212 228 BY SIMILARITY.  
FT DISULFID 337 353 BY SIMILARITY.  
FT DISULFID 364 393 BY SIMILARITY.  
FT CARBOHYD 144 144 N-LINKED (GLCNAC..). (POTENTIAL).  
SQ SEQUENCE 418 AA; 46263 MW; F4BC1DB020CFBBD0 CRC64;  
  
Query Match 37.78; Score 497.5; DB 1; Length 418;  
Best Local Similarity 40.5%; Pred. No. 1.3e-39;  
Matches 100; Conservative 44; Mismatches 82; Indels 21; Gaps 8;  
  
QY 1 VVGGRDADGEGMPQVSLHALGQGHICGASLISPNVLVSAAHGYIDDRGFYSDPTQWTA 60  
DB 187 ILGGTEAEGSPQVSLR-LNNAHCGGSLNNMWTAAHCFRSN-----SNPRDWIA 240  
QY 61 FLGLHDQSORAGPVQERRLRKRIISHFFNDFTDYDIALLEKPAEYSSMYRPICLPD 120  
DB 241 TSGI-----STTFPKLR-MVRNLIHNHNYSAIHENDIALVRLNLSVTFKDIHSCVCPA 295  
QY 121 ASHVFPAGKAIWVTGWHGTQYGGTGALTLQKEIRVINQITTCENLLPQQ-----ITPRMNC 176  
DB 296 ATQNTPPSTAYVTGWGAQAYAGHTVPELRGQVRIISNDVCN--APHSYNGAILSGMLC 353  
QY 177 VGLFSGVDSCGSGGSLSSVEADGR--IPQAGVSWGDCGCAORNPFGVYTRLPFRDW 234  
DB 354 AGVPQGVDAQCQDGGSGPL--VQEDSRRLMFIIVGWSGDCGLPDPGVYTRVAYLDW 411  
QY 235 IKENTGV 241  
DB 412 IRQQIGI 418  
  
RESULT 10  
TMS2\_HUMAN STANDARD; PRT; 492 AA.  
ID TMS2\_HUMAN STANDARD; PRT; 492 AA.  
AC O15393; Q9BXX1;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).  
GN TMPRSS2 OR PRSS10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97468144; PubMed=9325052;  
RA Paoloni-Giacobino A., Chen H., Peitsch M.C., Rossier C.,  
RA Antonarakis S.E.;  
RT "Cloning of the TMPRSS2 gene, which encodes a novel serine protease  
RT with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3";  
RL Genomics 44:309-320(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21309069; PubMed=11414763;  
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;  
RT "Mutation analyses of 268 candidate genes in human tumor cell lines";  
RL Genomics 74:352-364(2001).  
RN [3]  
RP SEQUENCE FROM N.A.; AND MUTAGENESIS.  
RX MEDLINE=21139112; PubMed=11245484;

RA Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Saffran D.C.,  
RA Raitano A.B., Jakobovits A.,  
RT "Catalytic cleavage of the androgen-regulated TMPRSS2 protease results  
in its secretion by prostate and prostate cancer epithelia.";  
RL Cancer Res. 61:1686-1692(2001).  
RN [4]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=21104370; PubMed=11169526;  
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;  
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and  
human tissues.";  
RL J. Pathol. 193:134-140(2001).  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY  
CLEAVAGE AND SECRETED.  
CC -!- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO  
EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U75329; AAC51784.1; -;  
DR EMBL: AF123453; RAD37117.1; -;  
DR EMBL: AF270487; AAK29280.1; -;  
DR EMBL: P00763; IDFO.  
DR MIM: 602060; -;  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR002172; LDL\_recept\_A.  
DR InterPro: IPR001190; SRCR.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00057; ldl\_recept\_a; 1.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: P00722; CHYMOTRYPSIN.  
DR SMART: SM00192; LDLA; 1.  
DR SMART: SM00202; SR; 1.  
DR SMART: SM00020; TRYP\_SPC; 1.  
DR PROSITE: PS01209; LDLRA\_1; 1.  
DR PROSITE: PS00068; LDLRA\_2; 1.  
DR PROSITE: PS00420; SRCR\_1; FALSE\_NEG.  
DR PROSITE: PS0287; SRCR\_2; 1.  
DR PROSITE: PS0240; TRYPsin\_DOM; 1.  
DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
DR Hydrolase: Serine protease; Transmembrane; Signal-anchor; Zymogen;  
KW Polymorphism.  
FT CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-  
FT CHAIN 256 492 CATALYTIC CHAIN.  
FT CHAIN 492 TRANSMEMBRANE PROTEASE, SERINE 2,  
FT CATALYTIC CHAIN.  
FT CYTOPLASMIC (POTENTIAL).  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT EXTRACELLULAR (POTENTIAL).  
FT LDL-RECEPTOR CLASS A.  
FT SRCR.  
FT SERINE PROTEASE.  
FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 345 345 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 441 441 CHARGE RELAY SYSTEM.  
FT SITE 255 256 CLEAVAGE (POTENTIAL).  
FT BY SIMILARITY.  
FT DISULFID 113 126 BY SIMILARITY.  
FT DISULFID 120 139 BY SIMILARITY.  
FT DISULFID 133 148 BY SIMILARITY.  
FT DISULFID 172 231 BY SIMILARITY.  
FT DISULFID 185 241 INTERCHAIN (BY SIMILARITY).

FT DISULFID 244 365 BY SIMILARITY.  
FT DISULFID 281 297 BY SIMILARITY.  
FT DISULFID 410 426 BY SIMILARITY.  
FT DISULFID 437 465 BY SIMILARITY.  
FT CARBOHYD 213 213 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 249 249 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT VARIANT 449 449 K -> N (IN DBSNP:1056602).  
FT /FTId=VAR\_011692.  
FT MUTAGEN 255 255 R->Q: LOSS OF CLEAVAGE.  
FT MUTAGEN 441 441 S->A: LOSS OF ACTIVITY.  
FT CONFLICT 160 160 M -> V (IN REF. 3).  
FT CONFLICT 242 242 I -> L (IN REF. 1).  
FT CONFLICT 329 329 E -> Q (IN REF. 1).  
FT CONFLICT 489 491 RAD -> KAN (IN REF. 1).  
FT SEQUENCE 492 AA: 53891 MW: CAB44FDI74A9076B CRC64;  
Query Match 37.7%; Score 497; DB 1; Length 492;  
Best Local Similarity 41.4%; Pred. No. 1.7e-39;  
Matches 99; Conservative 45; Mismatches 81; Indels 14; Gaps 7;  
QY 1 VVGTTDADGEMPMQVSLHALGQGHICGASLISPMWLYSAARHCYIDDRGFYSOPTOWTA 60  
Db 256 IYGGESALPGAMPQVSLH-VONVHVCSSITPEWIVTAHCVKEP-----LNNPWHWTA 310  
QY 61 FLGLHDQS-QRSAPGVQERLRKRIISHPFENDFTFDYDIALLELEKPAEYSSMYRPICLP 119  
Db 311 FAGILRQSPMEYAGYQ---VEKVISHPNDYDKTKNDIALMKLQKPLTFNDLVKPVCLP 367  
QY 120 DASHVFPAGKATWVTGWHTQYGGTGALILQKGEIRVINOVTCE--LLPQIITPRMVCV 177  
Db 368 NFGMMLQPEQLCWISGWGATEKGTSEVLNAKRVLLIETQRCNSRYVYDNLITPAMICA 427  
QY 178 GFLSGVSDCGSDSGPLSSVEADGRI-FQAGVSVSGDGCACRKNKPGVYTRLPFRMWI 235  
Db 428 GFLQGNVSDCGSDSGPL--VTSKNNIWLIGDTSWGSCKAKAYRPGVYGNVWVETDWI 484  
RESULT 11  
TMS5\_MOUSE  
ID TMS5\_MOUSE STANDARD; PRT; 455 AA.  
AC Q9ER04; Q9ER03; Q9ER02;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).  
GN TMPRSS5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
RC TISSUE=Brain;  
RA Mitsui S., Yamaguchi N.;  
RT "cDNA cloning of mouse spinesin.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 4).  
RC TISSUE=Brain;  
RA Mitsui S., Yamaguchi N.;  
RT "Molecular cloning of mouse type 4 spinesin.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (potential).  
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2, 3 AND 4 (SHOWN HERE);  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL; AB016229;	BAB20276.1;	-
DR	EMBL; AB016230;	BAB20277.1;	-
DR	EMBL; AB016423;	BAB20278.1;	-
DR	EMBL; AB041037;	BAB40328.1;	-
DR	MGI; MGI:1933407;	Tmprss5.	
DR	InterPro; IPR001314;	Chymotrypsin.	
DR	InterPro; IPR001190;	SRCR.	
DR	InterPro; IPR001254;	Trypsin.	
DR	Pfam; PF00089;	trypsin.1	
DR	PRINTS; PRO0722;	CHIMOTRYPSIN.	
DR	SMART; SMO0202;	SR; 1.	
DR	SMART; SMO0020;	Tryp_Spc; 1.	
DR	PROSITE; PS02040;	TRYPSIN_DOM; 1.	
DR	PROSITE; PS00134;	TRYPSIN_HIS; 1.	
DR	PROSITE; PS00135;	TRYPSIN_SER; FALSE_NEG.	
DR	PROSITE; PS00420;	SRCR_1; FALSE_NEG.	
DR	PROSITE; PS50287;	SRCR_2; 1.	
KW	Hydrolase; Serine protease;	Transmembrane; Signal-anchor;	
KW	Glycoprotein; Alternative splicing	CYTOSOLASMIC (POTENTIAL).	
FT	DOMAIN	1 49	(POTENTIAL).
FT	DOMAIN	50 70	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT	DOMAIN	71 455	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	112 207	SRCR.
FT	DOMAIN	218 455	SERINE PROTEASE.
FT	ACT_SITE	258 258	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	308 308	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	405 405	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	SITE	217 196	CLEAVAGE (POTENTIAL).
FT	DISULFID	135 196	BY SIMILARITY.
FT	DISULFID	148 206	BY SIMILARITY.
FT	DISULFID	209 328	BY SIMILARITY.
FT	DISULFID	243 259	BY SIMILARITY.
FT	DISULFID	374 390	BY SIMILARITY.
FT	DISULFID	401 429	BY SIMILARITY.
FT	CARBOHYD	163 163	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	170 170	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	319 319	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	375 375	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	1 144	MISSING (IN ISOFORM 2).
FT	VARSPLIC	1 10	MISSING (IN ISOFORM 3).
FT	VARSPLIC	1 182	MISSING (IN ISOFORM 1).
FT	VARSPLIC	183 192	GGVGEAWKP -> MEAQVGLLWV (IN ISOFORM 1).
FT	CONFLICT	325 325	D > G (IN REF. 1; BAB20277).
SQ	SEQUENCE	455 AA;	SCFC31789C6899AA SCRC64:

Query Match	36.6%	Score 483;	DB 1;	Length 455;
Best Local Similarity	40.6%;	Pred. No. 3.4e-38;		
Matches 99;	Conservative 36;	Mismatches 99;	Indels 10;	Gaps 16;
QY	1	VVGGTDADEGWPQCVSLHAGQCHIGCASLISPWNVLSAAHCYIDDRGFRYSDDPTOWTA	60	
Db	218	IVGGQAVASGRWPQASV-MLGSRHTCGASVLAPHVNVYTAACHMY--SFRLSLSSNRV	273	
QY	61	FLGLHDQQRSGAPGVQERRAKRIISHFFNFDDTFDVLALLEKPAEYSSMWRCILPD	120	
Db	274	HAGL--VSHGAVRQGHGTMEKIIPHLYSAQHNHDYDVALQLRTPINFSDTVDAVCLPA	331	
QY	121	ASHVFPAGKATVWTGWTGTQGGP-GALILQKEIRVINQTTCEN--LLPQGITPRMCMV	177	
Db	332	KEQYFPWGSQCVSWGHTDPSHT:HSSDT-LQDTMVLPLSTHLCNSSCMYSGALTHRLCA	391	
QY	178	GLSGGVSDCGDGGPLSSVEADGRIFQAGVYSGWGGCAQRNKGVYTRLPFLPDWIKKE	237	
Db	392	GILDGRAACDGGDGGPLVCPSSGP-TWHLVGVYSGWGGCAEPRNPGEVAKVAEFLDWIHD	450	
QY	238	NTGQ	241	

Db 451 TVOV 454

RESULT 12

STUB_DROME	STANDARD;	PRT;	786 AA.
ID	Q05319;		
AC	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	Serine proteinase stubble (EC 3.4.21.-) (Stubble-stubblod protein).		
GN	SB OR SB-SBD.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=OREGON-R;		
RX	MEDLINE=93281671; PubMed=7685111;		
RA	Appel L.F., Prout M., Abu-Shumays R., Hammonds A., Garbe J.C.,		
RA	Fristrom D., Fristrom J.;		
RT	"The Drosophila Stubble-stubblod gene encodes an apparent		
RT	transmembrane serine protease required for epithelial		
RT	morphogenesis.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 90:4937-4941(1993).		
CC	-!- FUNCTION: HORMONE DEPENDENT PROTEASE REQUIRED FOR EPITHELIAL		
CC	MORPHOGENESIS. HAS A DUAL FUNCTION, DETACHES IMAGINAL DISC CELLS		
CC	FROM EXTRACELLULAR MATRICES THROUGH ITS EXTRACELLULAR PROTEOLYTIC		
CC	DOMAIN AND TRANSMITS AN OUTSIDE-TO-INSIDE SIGNAL TO ITS		
CC	INTRACELLULAR DOMAIN TO MODIFY THE CYTOSKELETON DURING		
CC	MORPHOGENESIS. MAY BE ABLE TO ACTIVATE ITSELF.		
CC	-!- SUBCELLULAR LOCATION: TYPE II membrane protein.		
CC	-!- INDUCTION: BY 20-HYDROXYCYDSONE (20HE).		
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE		
CC	TRYPsin FAMILY.		
CC	-!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-24 IS THE INITIATOR.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC			
CC	EMBL: L11451; AAA28918.1; -		
DR	HSSP: P20231; IAAO.		
DR	DR MEROPS: S01.225; -		
DR	Flybase: FBgn003319; Sb.		
DR	InterPro: IPR001314; Chymotrypsin.		
DR	InterPro: IPR001254; Trypsin.		
DR	Fram: PFO089; trypsin; 1.		
DR	PRINTS: PR00722; CHYMOTRYPSIN.		
DR	SMART: SM00020; Tryp_Spc; 1.		
DR	PROSITE: PS02040; TRYPsin_DOM; 1.		
DR	PROSITE: PS00134; TRYPsin_HIS; FALSE_NEG.		
DR	PROSITE: PS00135; TRYPsin_SER; 1.		
KW	Hydrolase; Serine protease; Transmembrane; Glycoprotein; Zymogen;		
KW	Signal-anchor.		
FT	CHAIN 1 542		NON-CATALYTIC CHAIN (POTENTIAL).
FT	CHAIN 543 786		CATALYTIC CHAIN (POTENTIAL).
FT	DOMAIN 1 58		CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 59 80		SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT	DOMAIN 81 786		EXTRACELLULAR (POTENTIAL).
FT	DOMAIN 543 786		SERINE PROTEASE.
FT	DOMAIN 267 276		POLY-SER.
FT	DOMAIN 297 298		POLY-GLN.
FT	DOMAIN 381 478		SER/THR-RICH.
FT	DOMAIN 412 422		POLY-THR.
FT	DOMAIN 471 478		POLY-SER.
FT	DOMAIN 1 542		NON-CATALYTIC CHAIN (POTENTIAL).
FT	CHAIN 543 786		CATALYTIC CHAIN (POTENTIAL).
FT	DOMAIN 1 58		CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 59 80		SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT	DOMAIN 81 786		EXTRACELLULAR (POTENTIAL).
FT	DOMAIN 543 786		SERINE PROTEASE.
FT	DOMAIN 267 276		POLY-SER.
FT	DOMAIN 297 298		POLY-GLN.
FT	DOMAIN 381 478		SER/THR-RICH.
FT	DOMAIN 412 422		POLY-THR.
FT	DOMAIN 471 478		POLY-SER.



DR	HSSP; P00734; 2HNT.
DR	MEROPS; S01.224; .
DR	MGD; MGI:1196620; Hpn.
DR	InterPro; IPR001314; Chymotrypsin.
DR	InterPro; IPR001190; SCRC.
DR	InterPro; IPR001254; Trypsin.
DR	Fram: PF00089; trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	SMART; SM00202; SR; 1.
DR	SMART; SM00020; Tryp_SPC; 1.
DR	PROSITE; PS02040; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
DR	Hydrolase; Serine protease; Transmembrane; Signal-anchor.
KW	CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
FT	CHAIN 162 416 CATALYTIC CHAIN (POTENTIAL).
FT	DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT	DOMAIN 44 416 EXTRACELLULAR (POTENTIAL).
FT	DOMAIN 162 416 SERINE PROTEASE.
FT	ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 256 256 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 352 352 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DISELFD 152 276 INTERCHAIN (BY SIMILARITY).
FT	DISELFD 187 203 BY SIMILARITY.
FT	DISELFD 321 337 BY SIMILARITY.
FT	DISELFD 348 380 BY SIMILARITY.
FT	CARBOHYD 111 111 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ	SEQUENCE 416 AA; 44739 MW; 432194FF4004F848 CRC64;
 Query Match 36.5%; Score 481.5; DB 1; Length 416; Best Local Similarity 40.1%; Pred. No. 4, 2e-38; Matches 99; Conservative 41; Mismatches 86; Indels 21; Gaps	
QY	1 VYGDTDADEGWPQVSLHALGGGHGICGASLSPNWLVSAAHCYIDRGRYSDDPTQWTA 60 ::: : :  : :
Db	162 IVGGDGSSLGRWPQVSLRYDGT-HLCGGSLSGDWLVLTAAHCFPE-----RNRLVLSRWV 216 ::: : :  : :
QY	61 FLGLEHDQSORAPGVQVERLLKRIISH-----PFNEDETDF--YDIALLELEKPAEYSSMW 113 
Db	217 FAG---AVARTSPHAVQLVGQVAIVYHGVLPP-FRDPTIDENSNDIALVHLSSSLPTEYI 272 
QY	114 RTCLPDSASHVFPAKGALWVTGWGHTQYGGTGALILQKEIRVINQITCE--NLLPQQIT 171 ::: : :  : :
Db	273 QPVCLPARGAQLVDGKVCTVTGWGNTQFYGQQAAMVLEARYPIILSNVCNSPDFYGNQIK 332 ::: : :  : :
QY	172 PRMNCVGLSGVSDSCGDSGGPL----SSVEADGRIFAGVVGVNGDCCAQRNPGVYTRL 228  : :
Db	333 PRMFACAGYPEGIDACOGDSGGPFVEDSIGTSRWRLCGIVSWGTCALARKPGVYTKV 392  : :
QY	229 PLFRDWI 235   : :
Db	393 TDEFREWI 399
 RESULT 14 TMS5_HUMAN ID TMS5_HUMAN STANDARD; PRT: 457 AA. AC O9H3S3; DT 16-OCT-2001 (Rel. 40, Created) DT 16-OCT-2001 (Rel. 40, Last sequence update) DT 01-NAR-2002 (Rel. 41, Last annotation update) DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin). GN TMPSRS5. OS Homo sapiens (Human). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. OX NCBI_Taxid=9606; RN [1] RP SEQUENCE FROM N.A. RC TISSUE=Brain; RC	



DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;  
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;  
KW Repeat.  
FT SIGNAL 1 19 PLASMA KALLIKREIN HEAVY CHAIN.  
FT CHAIN 20 390 PLASMA KALLIKREIN LIGHT CHAIN.  
FT DOMAIN 391 638 APPLE 1.  
FT DOMAIN 110 195 APPLE 2.  
FT DOMAIN 200 285 APPLE 3.  
FT DOMAIN 291 376 APPLE 4.  
FT DOMAIN 389 621 SERINE PROTEASE.  
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 453 453 O-LINKED (PROBABLE).  
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT ACT\_SITE 434 434 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 483 483 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 578 578 CHARGE RELAY SYSTEM.  
FT DISULFID 21 104 BY SIMILARITY.  
FT DISULFID 47 77 BY SIMILARITY.  
FT DISULFID 51 57 BY SIMILARITY.  
FT DISULFID 111 194 BY SIMILARITY.  
FT DISULFID 137 166 BY SIMILARITY.  
FT DISULFID 141 147 BY SIMILARITY.  
FT DISULFID 201 284 BY SIMILARITY.  
FT DISULFID 227 256 BY SIMILARITY.  
FT DISULFID 231 237 BY SIMILARITY.  
FT DISULFID 292 375 BY SIMILARITY.  
FT DISULFID 318 347 BY SIMILARITY.  
FT DISULFID 322 328 BY SIMILARITY.  
FT DISULFID 340 345 BY SIMILARITY.  
FT DISULFID 383 503 BY SIMILARITY.  
FT DISULFID 419 435 BY SIMILARITY.  
FT DISULFID 517 584 BY SIMILARITY.  
FT DISULFID 548 563 BY SIMILARITY.  
FT DISULFID 574 602 BY SIMILARITY.  
SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;

Query Match 36.2%; Score 477; DB 1; Length 638;  
Best Local Similarity 39.3%; Pred. No. 1.8e-37;  
Matches 96; Conservative 43; Mismatches 91; Indels 14; Gaps 7;  
QY 1 VVGTDADGEWPMQVSLHA--LGQGHICGASLISPNWLVSAAHCYIDDRGRYSPTQW 58  
Db 391 IVGTNASLGEWPMQVSLQVSLVSTHLCGSGSIGROWLTAHCF---DGIPY--PDVW 445  
QY 59 TAFGLHDQSO--RSAPGVQERRLKRIISHFFNDFDYDIALLEKPAEYSSMVRPI 116  
Db 446 RIYGGILSLSEITKTEP---SSRIKELIIHQEVKVSBNYDIALIKLQPLNYPEKPI 502  
QY 117 CLPDASHVFPAGKAIWVGHTQYGGTGALILQKGEIRVINQTTGENTLLPOQ-ITPRM 175  
Db 503 CLPSKADNTIYTCWVTGWTGKTEQGETONILQKATIPVNEECCKYRDYVINKQMI 562  
QY 176 CVGFLSGGVDSOGDGGPLSSVEADGRIFQAGVSWGDCQAQRNKPQVYTRLPFRDWI 235  
Db 563 CAGYKEGTDACKDGGGPL-VCKHSGRWLVGITSWGEGCGRKGDPGVYTKVSEYMDWI 621  
QY 236 KENT 239  
Db 622 LEKT 625

Search completed: July 22, 2002, 17:12:43  
Job time: 128 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:11:05 ; Search time 26.51 Seconds  
(without alignments)  
1572.681 Million cell updates/sec

Title: US-09-657-986B-2  
Perfect score: 1319  
Sequence: 1 VVGGTDADEGEWPNQVSLHA.....PGVYTRLPFRDWIKENTGV 241

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL19.\*

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mnc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1151	87.3	855	11 Q9J17	Q9J17 rattus norv
2	865	55.6	845	13 Q9DGR1	Q9DGR1 xenopus lae
3	585.5	44.4	799	11 Q9DB10	Q9DB10 mus musculus
4	524.5	39.8	279	11 Q9QZ74	Q9QZ74 rattus norv
5	509	38.6	767	13 Q9DGR2	Q9DGR2 xenopus lae
6	501.5	38.0	250	5 Q9V514	Q9V514 drosophila
7	500	37.9	490	11 Q920K3	Q920K3 rattus norv
8	497	37.7	492	4 Q96T73	Q96T73 homo sapien
9	489	37.1	389	13 Q9PVX7	Q9PVX7 xenopus lae
10	483	36.6	787	5 Q9VER6	Q9VER6 drosophila
11	478.5	36.3	624	11 Q9DAT3	Q9DAT3 mus musculus
12	474.5	36.0	624	11 Q91Y47	Q91Y47 mus musculus
13	463.5	35.1	1524	13 Q91674	Q91674 xenopus lae
14	463	35.1	643	6 Q97506	Q97506 sus scrofa
15	462	35.0	329	13 Q42272	Q42272 xenopus lae
16	461	35.0	274	5 O16133	O16133 anopheles s

#### ALIGNMENTS

RESULT 1  
Q9J17 ID Q9J17 PRELIMINARY; PRT; 855 AA.  
AC Q9J17;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE MEMBRANE BOUND SERINE PROTEASE (MEMBRANE BOUND ARGININE SPECIFIC  
DE SERINE PROTEASE).  
GN MSP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WISTAR; TISSUE=JEJUNUM;  
RA Tsuzuki S.;  
RT "A membrane bound serine protease expressed in rat small intestine.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WISTAR; TISSUE=DUODENUM;  
RA Iroue H., Takahashi K., Kishi K.;  
RT "membrane-bound arginine specific serine protease.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
DR EMBL; AB037898; BAB03502.1; .  
DR EMBL; AB049189; BAB13765.1; .  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.302; .  
DR InterPro; IPR002106; AA\_rnase\_ligase\_II.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR002172; LDL\_recept\_A.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00431; CUB; 2.

17	461	35.0	537	4 Q9BYE1	Q9BYE1 homo sapien
18	461	35.0	581	4 Q9BYE2	Q9BYE2 homo sapien
19	455.5	34.5	812	11 Q91WJ5	Q91WJ5 mus musculus
20	453	34.3	258	5 Q97399	Q97399 phaeton coc
21	453	34.3	334	6 Q46507	Q46507 papio hamad
22	453	34.3	810	4 Q15146	Q15146 homo sapien
23	449	34.0	624	6 Q35WE7	Q35WE7 oryctolagus
24	448.5	34.0	263	5 O02570	O02570 culex quinq
25	447.5	33.9	263	5 O62562	O62562 penaeus van
26	446	33.8	339	11 Q99L44	Q99L44 mus musculus
27	445	33.7	274	5 Q17086	Q17086 anopheles s
28	444.5	33.7	812	11 Q9R0W3	Q9R0W3 rattus norv
29	444	33.7	454	6 Q46506	Q46506 papio hamad
30	442.5	33.5	317	13 Q9DGR3	Q9DGR3 xenopus lae
31	442	33.5	261	13 Q9W7Q4	Q9W7Q4 paralichthy
32	442	33.5	273	6 Q9XSM1	Q9XSM1 ovis aries
33	442	33.5	364	5 Q917V4	Q917V4 drosophila
34	436.5	33.1	467	5 Q967X8	Q967X8 panulirus a
35	435.5	33.0	297	11 Q88781	Q88781 rattus ratt
36	435	33.0	263	11 Q9CR35	Q9CR35 mus musculus
37	433.5	32.9	461	6 Q95ND7	Q95ND7 pan troglod
38	433	32.8	242	13 Q9W7Q7	Q9W7Q7 paralichthy
39	433	32.8	405	4 Q96E86	Q96E86 homo sapien
40	432	32.8	263	11 Q9DC86	Q9DC86 mus musculus
41	431	32.7	273	11 Q921N4	Q921N4 mus musculus
42	429.5	32.6	310	11 Q9QYZ9	Q9QYZ9 mus musculus
43	429.5	32.6	456	4 Q14316	Q14316 homo sapien
44	429	32.5	242	13 Q92099	Q92099 parantotho
45	429	32.5	263	11 Q9D8X8	Q9D8X8 mus musculus

```

DR      InterPro: IPR0021172; LDL_recept_A.
DR      InterPro: IPR001254; Trypsin.
DR      Pfam: PF00431; CUB: 2.
DR      Pfam: PF00057; ldl_recept_a; 4.
DR      Pfam: PF00089; trypsin; 1.
DR      PRINTS: PR00722; CHYMOTRYPSIN.
DR      PRINTS: PR00261; LDLRECEPTOR.
DR      SMART: SM00042; CUB: 2.
DR      SMART: SM00192; LDLa; 3.
DR      SMART: SM00020; Tryp_SPC; 1.
DR      PROSITE: PS01180; CUB: 2.
DR      PROSITE: PS00020; LDLa_2; 4.
DR      PROSITE: PS00240; TRYPSIN_DOM; 1.
DR      PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR      GlycoProtein: Hydrolase; Serine protease.
KW      SEQUENCE 845 AA; 93597 MW; 7FD7E62851A758B7 CRC64;
SQ

```

Query Match	65.6%	Score	865.	DB	13	Length	845;
Best Local Similarity	62.12%	Pred. No.	2.1e-78;				
Matches	150;	Conservative	41;	Mismatches	50;	Indels	0;
Qy	1	VVGTDADGEWPQVSLHALQGQHICGASLISPNWLVSAAHYIIDRGRFYSDPTQMTA	60				
Db	605	IVGVNADTGEFPQVSLHAKGNKKTCCASLGFFMTLISAACHQDDHQMYSDSALMTA	664				
Qy	61	PLGLHDSQSAPGVQERLKLRIISHPFNFDFDYDIALLELEKPAEYSSMVRPCLPD	120				
Db	665	YLGHLHQAQLNTKVVVEREIKRIMAHIGFNONTYDNDIAVLELEKPEYVTDFIQPVCIPE	724				
Qy	121	ASHVFPAKRAIWTGWGHTQYQGTGALILQKGEIRVINQTTCLNLPQQITPRMNCVGL	180				
Db	725	STHDFPVGKPIWYTGWALKKEGGAAVILQKAEIRIINQTECNKLLDGLTPRMLCAGEV	784				
Qy	181	SGGVDSQCGDSGGPLSSVREADGRITFQAGVWSGDCQARNKPGVYTRPLFRDRIKXGNTG	240				
Db	785	SGGIDACQDGGSGPLSSVELNKKVYLAGVWSGEGCARRNKPGVYFKVSMKRDWSKDKTG	844				
Qy	241	V	241				
Db	845	L	845				

```

RESULT      3
Q9DB10      PRELIMINARY;      PRF;      799 AA.
ID Q9DB10
AC Q9DB10;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE D1300008A22RIK PROTEIN.
GN 1300008A22RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER.
RX MEDLINE=21083660; PubMed=11217851;
RA Kawai J., Shiragadawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukuishii Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bozell T., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

```

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
CC EMBL; AK004939; BAB23684.1; -.  
CC HSSP; P00763; IDPO.  
CC MGD; MGI:1919003; 1300008A22Rik.  
CC InterPro; IPR001314; Chymotrypsin.  
CC InterPro; IPR000859; CUB.  
CC InterPro; IPR002172; LDL\_recept\_A.  
CC InterPro; IPR001254; Trypsin.  
CC Pfam; PF00087; LDL\_recept\_a; 3.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC SMART; PR00261; LDLRECEPTOR.  
CC SMART; SM00042; CUB; 1.  
CC SMART; SM00192; LDLa; 3.  
CC SMART; SM00020; Tryp\_SPC; 1.  
CC PROSITE; PS01180; CUB; 1.  
CC PROSITE; PS00068; LDLRA\_2; 3.  
CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
KW Glycoprotein; Hydrolase; Serine protease.  
SQ SEQUENCE 799 AA; 89557 MW; 16315A646AD5288 CRC64;

Query Match 44.4%; Score 585.5; DB 11; Length 799;  
Best Local Similarity 45.6%; Pred. No. 2.9e-50;  
Matches 108; Conservative 43; Mismatches 81; Indels 5; Gaps 3;

QY 1 VVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGRYSDDPTQWT 60  
DB 565 IVGTVSSEGEWPMQVSLQIRG-HICGGLIADRWVITAAHCFOED---SNASPKLTV 620  
QY 61 FLGLHDQSQRSAPGVQERRLRKRIISHPPFNDFDYLALLELEKPAEYSSMVRPICLPD 120  
DB 621 FLGKMRQNSR-WPGEVSEKVSRLFLHPYHEEDSHDVALQLDHPVYSATVRPVCLPA 679  
QY 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINTOTTENLLPQITPRM 180  
DB 680 RSHFFEPQGHCHWITGWAQREGPVSTLQKVDVQLVPODLCEAYRYQVSPMLCAGYR 739  
QY 181 SGGVDSGCGSGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYTRLPFRDWIKE 237  
DB 740 KKKKDACGSGDGLVCEPSGRWFLAGLVSWGLGCRPNFPGVYTRVTRVNNIQ 796

RESULT 4  
Q90274 PRELIMINARY; PRT; 279 AA.  
AC Q90274;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE ADRENAL SECRETORY SERINE PROTEASE PRECURSOR.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NEDE;  
RX MEDLINE=21332587; PubMed=11439186;  
RA Bicknell A.B., Lomthasong K., Woods R.J., Hutchinson E.G.,  
RA Bennett H.P.J., Gladwell R.T., Lowry P.J.;  
RT "Characterization of a Serine Protease that Cleaves Pro-gamma-  
RT Melanotropin at the Adrenal to Stimulate Growth.";

RL Cell 105:903-912(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC EMBL; AF198087; AAF13253.1; -.  
CC HSSP; P00760; IAQ7.  
CC MEROPS; S01.047; -.  
CC InterPro; IPR001314; Chymotrypsin.  
CC InterPro; IPR001254; Trypsin.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC SMART; SM00020; Tryp\_SPC; 1.  
CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease; Signal.  
RT SIGNAL 1.  
RT SIGNAL 22.  
SQ SEQUENCE 279 AA; 30522 MW; 1C28069DF0064546 CRC64;

Query Match 39.8%; Score 524.5; DB 11; Length 279;  
Best Local Similarity 41.5%; Pred. No. 1e-44;  
Matches 102; Conservative 42; Mismatches 83; Indels 19; Gaps 6;

QY 1 VVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGRYSDDPTQWT 59  
DB 48 IIGTQAEITGDWPMQVSLQ-LNNVHCGGTLISNLWYLAHC-----FRSYSNPQWT 100  
QY 60 AFLGLHDQSQRSAPGVQERRLRKRIISHPPFNDFDYLALLELEKPAEYSSMVRPICLP 119  
DB 101 ATFGVSTISR-----LRVVRALLAHAEYNSITRDNDIAVQVLDPRVTFTRIHRVCLP 155  
QY 120 DASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINTOTTENLLPQITPRM 175  
DB 156 AATQNIIPDSVAVATGWGSLTYGNTVTNLQGEVIRYSVGVNE--PAGYGGSVLPGLM 213  
QY 176 CVGFLSGVDSGCGSGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYTRLPFRDWI 235  
DB 214 CAGVRSAGVACQGDGSGPLVQEDTRLENFVVGIVSWGQGLPNKPGVYTRVTRVNNI 273  
QY 236 KENTGV 241  
DB 274 RQQTGI 279

RESULT 5  
Q9DGR2 PRELIMINARY; PRT; 767 AA.  
AC Q9DGR2;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE EMBRYONIC SERINE PROTEASE-2.  
GN XESP-2.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20363741; PubMed=10903452;  
RA Yamada K., Takabatake T., Takeshima K.;  
RT "Isolation and characterization of three novel serine protease genes  
RT from Xenopus laevis.";  
RL Gene 252:209-216(2000).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC EMBL; AB038497; BAB08217.1; -.  
CC HSSP; P00766; ICHG.  
CC MEROPS; S01.049; -.  
CC InterPro; IPR001314; Chymotrypsin.  
CC InterPro; IPR002172; LDL\_recept\_A.  
CC InterPro; IPR001190; SRCR.



OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tsuzuki S.;  
RT TMRPSS2; Rat.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB073550; BAB70683.1; -;  
SQ SEQUENCE 490 AA; 53518 MW; 2BC691551CAC409A CRC64;

Query Match 37.9%; Score 500; DB 11; Length 490;  
Best Local Similarity 40.1%; Pred. No. 6.4e-42;  
Matches 99; Conservative 48; Mismatches 84; Indels 10; Gaps 5;

QY 1 VVGTDADGEWQVSLHALGOGHICGASLISPNMLVSAACHYIDDRGFYSDPTQWTA 60  
DB 1 VVGTDADGEWQVSLHALGOGHICGASLISPNMLVSAACHYIDDRGFYSDPTQWTA 60  
DB 254 IVGGTASPDGWPQVSLVHOGI-HVCGGSIITPEWIVTAREHVEEP-----LSSPRYWTA 308  
QY 61 FLGLHDQSQRSAFQVQERLKRRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPCLPD 120  
DB 309 FAGILKOSLMFYG--SRHQVEKVISHPNYDSKTKNNDIALMKLQTLPLAFNDVVKVCLPN 366  
QY 121 ASHVFPAGKAIWGTGHTQYGTGALILQKGEIRVINGTCEN--LLPQQTIPRMVCGV 178  
DB 367 PGMLDLAQECWISGWGATYKGTSDVLNAAMVPLIEPSKNSKYIYNLLITPAMICAG 426  
QY 179 FLSCGVDSQCGSGGLSSVEADGRIFQAGVYVSWGDCGAQRNKPQVYTRLPFRDWI 235  
DB 427 FLOGVDSQCGSGGLPLVLKNE-IWNLIGDTSWGSGCAKAYRPGVYGNVTFTDWI 482

RESULT 8  
Q96T73 PRELIMINARY; PRT; 492 AA.  
AC Q96T73; 2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DE EPIHELIALIN.  
GN TMRPSS2  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2123025; PubMed=11322890;  
RA Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,  
RA Hoidal J.R.;  
RT "Cloning and characterization of the cDNA and gene for human  
epithelialin.";  
RL Eur. J. Biochem. 268:2687-2699(2001).  
DR EMBL; AF329454; AAK53559.1; -;  
SQ SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DADF CRC64;

Query Match 37.7%; Score 497; DB 4; Length 492;  
Best Local Similarity 41.4%; Pred. No. 1.3e-41;  
Matches 99; Conservative 45; Mismatches 81; Indels 14; Gaps 7;

QY 1 VVGTDADGEWQVSLHALGOGHICGASLISPNMLVSAACHYIDDRGFYSDPTQWTA 60  
DB 256 IVGGESALPGAWPQVSLH-VQNVHVCVGGSIITPEWIVTAREHVEEP-----LNNPWHWTA 310  
QY 61 FLGLHDQS-QRSAPGVQERLKRRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPCLP 119  
DB 311 FAGILRQSFMEYAGYQ--VEKVISHPNYDSKTKNNDIALMKLQPLFNDLVKPVCLP 367  
QY 120 DASHVFPAGKAIWGTGHTQYGTGALILQKGEIRVINGTCEN--LLPQQTIPRMVCGV 177  
DB 368 NPGMMLQPEQLCWISGWGATEERKGTSEVNLNAKVLIIETQRCSNRYVDNLLITPAMICA 427

QY 178 GFUSGGVDSQCGSGGLSSVEADGRIFQAGVYVSWGDCGAQRNKPQVYTRLPFRDWI 235  
DB 428 GFLOGVDSQCGSGGL--VTSKNINWNLIGDTSWGSGCAKAYRPGVYGNVTFTDWI 484

RESULT 9  
Q9PVX7 PRELIMINARY; PRT; 389 AA.  
AC Q9PVX7;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE EPIDERMIS SPECIFIC SERINE PROTEASE.  
GN XEPSIN.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
OX Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K.;  
RT "The expression control of xepsin by non-axial and planar  
posteriorizing signals in Xenopus epidermis.";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
DR EMBL; AS018694; BAA84941.1; -;  
DR HSP; P00763; IDPO.  
DR MEROPS; S01.00A; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp-SPC; 1.  
DR PROSITE; PS02440; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 389 AA; 42375 MW; B31FB4A2F5D1F6E3 CRC64;

Query Match 37.1%; Score 489; DB 13; Length 389;  
Best Local Similarity 40.6%; Pred. No. 6e-41;  
Matches 102; Conservative 39; Mismatches 86; Indels 24; Gaps 8;

QY 1 VVGTDADGEWQVSLHALGOGHICGASLISPNMLVSAACHYIDDRGFYSDPTQWTA 60  
DB 26 IVGNDSKRGWNPQISL-SYKSDSICGSLDTSWVTAHC-IDS-----LDVSYIV 78  
QY 61 FLGLHDQSQRSAF--GVQERLKRRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPCL 118  
DB 79 YLGAY--QLSAPDNSTVSRGKSIITKHPDFQVEGSGDIALIELEKPVTFYILPCL 135  
QY 119 PDASHVFPAGKAIWGTGHTQYGG--TGALILQKGEIRVINGTCENLLPQ----- 169  
DB 136 PSQDVQPAAGTMCWVTGWGNIQEGTPLISFKTIQKAEVAIIDSSVCGTWYESSLYIPDF 195  
QY 170 --ITPRMVCVFLSGVDSQCGSGGLSSVEADGRIFQAGVYVSWGDCGAQRNKPQVYTR 227  
DB 196 SFIQEDMVCAGYKGRIDACQCGSGGPL-VCNVNWNVLQGLIVSWGYGCAEPNRPVYTK 254  
QY 228 LPLFRDWIKEN 238  
DB 255 VQYVQDWLKTN 265

RESULT 10  
Q9VEY6 PRELIMINARY; PRT; 787 AA.  
ID Q9VEY6  
AC Q9VEY6;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)



```

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SB GENE PRODUCT.
GN SB OR CG4316.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20195006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pachl J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Zhu S., Zhu X., Smith H.O.,
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC EMBL; AF003712; AAF55277.1; -.
CC HSSP; P00763; LDPO.
CC MEROPS; S01.225; -.
CC FlyBase; FBgn0003319; SB.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR002965; P_Rich_extensn.
CC InterPro; IPR001254; Trypsin.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPsin.
CC PROSITE; PS01217; PRICHEXTENSIN.
CC SMART; SM00020; TRYP_SPC; 1.
CC PROSITE; PS02040; TRYPsin_DOM; 1.
CC PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 787 AA; 85143 MW; 02B286750BD6025A CRC64;

Query Match 36.6%; Score 483; DB 5; Length 787;
Best Local Similarity 39.2%; Pred. No. 6,1e-40;
Matches 98; Conservative 47; Mismatches 85; Indels 20; Gaps 8;

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:08:39 ; Search time 13.06 Seconds  
(without alignments)  
450.733 Million cell updates/sec

Title: US-09-657-986B-2

Perfect score: 1319  
Sequence: 1 VVGSTDADEGEWPMQVSLHA.....PGVYTRLPFLFRDWIKENTGV 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	855	2	US-09-027-337-2
2	587.5	41.5	235	4	US-08-944-483-65
3	544.5	41.3	235	3	US-08-807-151-3
4	544.5	41.3	235	4	US-09-478-957-3
5	544.5	41.3	798	1	US-08-200-900A-2
6	544.5	41.3	798	5	PCT-US94-00616-2
7	497.5	37.7	232	1	US-08-508-448C-19
8	497.5	37.7	418	1	US-08-508-448C-25
9	497	37.7	482	4	US-08-342-749-2
10	488	37.0	283	3	US-08-807-151-1
11	488	37.0	283	4	US-09-478-957-1
12	485	36.8	454	4	US-09-518-046-2
13	481.5	36.5	415	2	US-09-000-846-2
14	473.5	35.9	255	4	US-08-944-483-67
15	473.5	35.9	256	2	US-09-027-337-3
16	469	35.6	638	2	US-08-661-151-3
17	462	35.0	248	4	US-08-944-483-63
18	461	35.0	299	4	US-08-944-483-66
19	455	34.5	238	4	US-08-944-483-64
20	453.5	34.4	812	1	US-08-429-743-1
21	453.5	34.4	812	1	US-08-451-932-1
22	453.5	34.4	812	1	US-08-452-260-1
23	453.5	34.4	812	2	US-08-612-788-1
24	453.5	34.4	812	2	US-08-605-598B-1
25	453.5	34.4	812	2	US-08-429-743-1
26	453.5	34.4	812	2	US-08-866-735-1
27	453.5	34.4	812	2	US-08-866-735-1

28	453.5	34.4	812	3	US-09-066-028-1	Sequence 1, Appli
29	453.5	34.4	812	5	PCT-US95-05107-1	Sequence 1, Appli
30	453	34.3	791	1	US-08-843-219-1	Sequence 1, Appli
31	453	34.3	791	3	US-08-851-350-1	Sequence 47, Appli
32	452	34.3	230	1	US-08-456-840-47	Sequence 47, Appli
33	452	34.3	230	1	US-08-266-407A-47	Sequence 47, Appli
34	452	34.3	230	2	US-08-892-544-47	Sequence 47, Appli
35	452	34.3	230	2	US-08-766-983-12	Sequence 12, Appli
36	452	34.3	230	4	US-08-944-483-53	Sequence 53, Appli
37	452	34.3	230	4	US-09-296-219-12	Sequence 12, Appli
38	452	34.3	546	6	5200340-6	Patent No. 5200340
39	452	34.3	790	1	US-08-469-486-54	Sequence 54, Appli
40	452	34.3	790	2	US-08-469-658-54	Sequence 54, Appli
41	452	34.3	791	2	US-09-131-995-1	Sequence 1, Appli
42	452	34.3	791	2	US-08-832-087B-1	Sequence 1, Appli
43	452	34.3	791	4	US-09-132-154-1	Sequence 1, Appli
44	452	34.3	810	1	US-07-854-603-2	Sequence 2, Appli
45	452	34.3	810	1	US-08-147-000B-29	Sequence 29, Appli

ALIGNMENTS

RESULT 1  
US-09-027-337-2  
; Sequence 2, Application US/09027337B  
; Patent No. 5972616

; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Tanimoto, Hiroto  
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in  
; TITLE OF INVENTION: Breast and Ovarian Carcinomas  
; FILE REFERENCE: D6064  
; CURRENT APPLICATION NUMBER: US/09/027,337B  
; CURRENT FILING DATE: 1998-02-20  
; NUMBER OF SEQ ID NOS: 13  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides  
; OTHER INFORMATION: 23 to 2589 of Sequence 1  
; Patent No. 5972616  
US-09-027-337-2

Query Match	100.0%	Score 1319;	DB 2;	Length 855;
Best Local Similarity	100.0%;	Pred. No. 1.5e-130;		
Matches 241;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	VVGSTDADEGEWPMQVSLHA		Sequence 1, Appli
Db	615	VVGSTDADEGEWPMQVSLHA		Sequence 1, Appli
QY	61	FLGLHDQSQRSAPGVQERRLKR		Sequence 1, Appli
Db	675	FLGLHDQSQRSAPGVQERRLKR		Sequence 1, Appli
QY	121	ASHVFPAGKAIWVGWCHTQYGGT		Sequence 1, Appli
Db	735	ASHVFPAGKAIWVGWCHTQYGGT		Sequence 1, Appli
QY	181	SGGVDSQCGSGGGLSSVEADGR		Sequence 1, Appli
Db	795	SGGVDSQCGSGGGLSSVEADGR		Sequence 1, Appli
QY	241	V 241		Sequence 29, Appli
Db	855	V 855		Sequence 29, Appli

Copyright



[illegible]

```

; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00616-2

```

[illegible][illegible]

Query Match	37.7%	Score 497.5;	DB 1;	Length 418;
Best Local Similarity	40.5%	Pred. No. 2.3e-44;		



Matches	100;	Conservative	44;	Mismatches	82;	Indels	21;	Gaps	8;
---------	------	--------------	-----	------------	-----	--------	-----	------	----

  

QY	1	VYGGTDADGEWPMQVSLHGLQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTA	60
		:	:
Db	187	ILGGTEABEGSWPMQVSLR-LNNAHCGGSLINNWMILTAAHCRSN-----SNPRDWT	240
		:	:
QY	61	FUGLHDQSORGAPGVOERLKRILISHPFNDFDXYDIALLELEKFAEYSSMVRPICLPD	120
		:	:
Db	241	TSGI-----STTFPLR-MRVNRILIHNNYSATHENDIALVRLENSVTTKDLSHVCPLA	295
		:	:
QY	121	ASHVPPACKAIWVTGWGHTGYGGTCALILQKGEIRVINQTTCCNLLPQO-----ITPRMMC	176
		:	:
Db	296	ATONIPPGSTAYVTGWAQAEVAGHTVPELRQGVRIISNDVCN--APHSVNGAILSGMLC	353
		:	:
QY	177	VGFSLGGVSDCGSDGGPLSSVEADGR--IFQAGVSWGDCGCAQRNKPQVYTRLPFLFRDW	234
		:	:
Db	354	AGVPOGGVDACOGDSGGPL--VQEDSRRLFWIFGIVSWGDCGLPDKPGVYTRVAYLDW	411
		:	:
QY	235	IKENTGV	241
		:	:
Db	412	IRQQTGI	418
		:	:

  

RESULT	9
US-09-342-749-2	
Sequence 2, Application US/09342749	
Patent No. 6166194	
GENERAL INFORMATION:	
APPLICANT: Wong, Alexander K.C.	
APPLICANT: Tavtigian, Sean V.	
APPLICANT: Teng, David H.-F.	
APPLICANT: Myriad Genetics, Inc.	
TITLE OF INVENTION: TMRPS2 is a Tumor Suppressor	
FILE REFERENCE: 2318-202	
CURRENT APPLICATION NUMBER: US/09/342,749	
CURRENT FILING DATE: 1998-06-29	
EARLIER APPLICATION NUMBER: US 60/091,044	
EARLIER FILING DATE: 1998-06-29	
NUMBER OF SEQ ID NOS: 33	
SOFTWARE: PatentIn Ver. 2.0	
SEQ ID NO 2	
LENGTH: 492	
TYPE: PRT	
ORGANISM: Homo sapiens	
US-09-342-749-2	

  

Query Match	37.7%;	Score	497;	DB	4;	Length	492;
Best Local Similarity	41.4%;	Pred.	No. 3.2e-44;				
Matches	99;	Conservative	45;	Mismatches	81;	Indels	14;
Gaps	7;						

  

QY	1	VYGGTDADGEWPMQVSLHGLQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTA	60
		:	:
Db	256	IVGGSALPGAMPQVSLH-VQNHVCGSIIITPEWIVTAAHCKEP---LNNPWHWTA	310
		:	:
QY	61	FUGLHDQS-QRSAPGVQERRLKRILISHPFNDFDXYDIALLELEKFAEYSSMVRPICLP	119
		:	:
Db	311	FAGILRQSPMFYAGAYQ---VEKVISHENYDSKTKNNIDIALMKLQKPLTFNDILVKVCVLP	367
		:	:
QY	120	DASHVPPACKAIWVTGWGHTGYGGTCALILQKGEIRVINQTTCCN-----LLPQQITPRMMC	177
		:	:
Db	368	NFGMMLQPLQCLWISGWGATEKGTSEVLNAKVLIIETQRCNSRVYDNLITPAKICA	427
		:	:
QY	178	GFUSGVDVSCQDSGGPLSSVEADGRI--FQAGVSWGDCGCAQRNKPQVYTRLPFLFRDWI	235
		:	:
Db	428	GFUGGVDVSCQDSGGPL--VTSKNIWILIGDT-SWGSGCAKAYRPGVYCNVAVFTDWI	484
		:	:

```

; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 133
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
US-09-518-046-2

Query Match 36.8%; Score 485; DB 4; Length 454;
Best Local Similarity 40.4%; Pred. No. 5,3e-43;
Matches 97; Conservative 44; Mismatches 85; Indels 14; Gaps

QY 1 VVGCTDADGEWPQVSLHALGQCHICGASLSPNWLYSAAHCYIDDRGPRYSDDPTQWTA 60
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 218 IIVGNNLSLSQMPQASLQFGY-HLGGGSVITPLWITAAHCYVD-----LYLPKSWTI 271
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 61 FLGLHDQSORGAPGVQERRKRIISHPFNFDETFDYDIALLELEKPAEYSVMVRPGLPD 120
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 272 QVGLVSLLDNFP---SHLVEKIVYHSKYKPKRLGNDIALMKLAGSLFTFNEMIQVCLPN 328

QY 121 ASHVFPACKATWVKGWHTQVGGTGALILQKGEIRVINGTTC--ENLLPQGITPRMCMVG 178
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 329 SEENFPDGKVCWTSGWATEDGGDASPLVNLHAAPLISNKNICNHDVYVGGIISPSMICAG 388
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 179 FLTGGVDSCQSDSGPSSVEADGRIFQ-AGVYVSGDGCQAQRNKPQGVYTRLPFLFRDWIKE 237
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 389 YLTGGVDSQDSDGGLP--VCQERLWLKLVGATSFQIGCAEVNKPQGVYTRVTSFLDWIHE 446

RESULT 13
US-09-000-846-2
; Sequence 2, Application US/09000846
; Patent No. 5981830

```

APPLICANT: SADLER, JASPER  
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH  
TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSES: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.  
STREET: 2200 CLARENDON BLVD. SUITE 1400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: US  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/000,846  
FILING DATE: 30-DEC-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/866,058  
FILING DATE: 30-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: LEOVITZ, RICHARD M.  
REGISTRATION NUMBER: 37,067  
REFERENCE/DOCKET NUMBER: BERLX 65P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-243-6333  
TELEFAX: 703-243-6410  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
TYPE: amino acid



Db 114 PVCLPAAGQALVDGKICTVTGNGTQYGGQAGVLQEARVPIISNDVCNGADFYGNQIKP 173  
Qy 173 RMCVGFSLGGVDSGCGDGGPL---SSVEADGRIFQAGVWSWGDGCAORNKPGVYTRLP 229  
Db 174 KMFCAGYPEGGIDACGDSGGPFVCEDSISRTPRWRLCGIVSWGTGCALAOXPGVYTKVS 233  
Qy 230 LFRDWI 235  
Db 234 DFREWI 239

Search completed: July 22, 2002, 17:12:25  
Job time: 226 sec